A COMPARISON OF TWO REVIEW METHODS FOR ALGEBRA AND TRIGONOMETRY AT WISCONSIN STATE UNIVERSITY, LACROSSE

By

ROLAND WALTER CHRISTENSEN

Bachelor of Arts

Iowa State Teachers College

Cedar Falls, Iowa

Submitted to the faculty of the Graduate School
of the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF EDUCATION
May, 1967

A COMPARISON OF TWO REVIEW METHODS FOR ALGEBRA
AND TRIGONOMETRY AT WISCONSIN STATE
UNIVERSITY, LACROSSE

Thesis Approved:

W Ware Marsden
Thesis Adviser

Milton E Berg

Gerald K. Goff

Vernon Troxel

D D Durham
Dean of the Graduate School

658405

ACKNOWLEDGEMENTS

The writer wishes to express his gratitude to all persons who contributed to the completion of this study. He especially wishes to express appreciation and gratitude to Dr. W. Ware Marsden, Chairman of the Advisory Committee, for his encouragement and guidance during this research project; to Dr. Milton Berg, Dr. Gerald Goff, and Dr. Vernon Troxel, members of the Advisory Committee, for their interest and assistance; and to Dr. Robert E. Sweitzer, Dr. Jeanne L. Agnew, Dr. John L. Folks, and Mrs. Helen M. Jones who gave inspiration and direction at the beginning of the research.

The cooperation of the Mathematics Department, the Audio-Visual Department, the administration and the students of Wisconsin State University, LaCrosse, who participated in this study is deeply appreciated. The writer is especially grateful to Mr. V. B. Rasmussen, Director of the Audio-Visual Center at Wisconsin State University, at LaCrosse, and Dr. M. O. Graff, Vice-President for Academic Affairs at Wisconsin State University, LaCrosse.

The writer expresses special gratitude to his wife, Laura, and to their sons, Steven, Andrew, and David, for their encouragement, patience, work and sacrifice.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page

CHAPTER I

INTRODUCTION

American mathematics education has been marked by a massive examination and re-examination, frequently agonizing and critical, during the past decade. This scrutiny was given much of its original impetus by the launching of Russia's first Sputnik. American scientists generally attributed this spectacular success to the superiority of Russian scientific education. The promptings of many well-known American scientists were the beginnings of a very critical self-examination by American science educators.

This scrutiny of mathematics education was originally directed toward its subject matter. This close examination, quite appropriately, concerned itself with the content and method of scholastic mathematics education. From this initial critical observation evolved the School Mathematics Study Group, the Ball State Program, and other similar programs.

While the original impetus for this examination of American mathematics education was directed towards an investigation of the subject matter of scholastic and collegiate mathematics, much of the more recent inspection of higher mathematics education has been directed towards the possible methods of teaching large sized classes of mathematics students in colleges and universities. Soaring enrollments prompted this inquiry.

The investigation herein reported was undertaken in recognition of the need for more information concerning feasible methods for handling large sized collegiate mathematics classes. Recognition was given to the fact that similar investigations have been and are being conducted. Every college and university curriculum has unique features. These features were deemed sufficiently numerous and unique to justify further examination. Since there has also been little research done on this problem in state supported institutions comparable to Wisconsin State University at LaCrosse, an investigation of this type is desirable and appropriate.

The basic question to be examined by this investigation is "Will the level of achievement of pre-calculus mathematics students be affected by the use of a type of mechanical teaching device?" The particular type of mechanical device will be described later in this essay. The type of device is one that seems particularly well adapted to an institution of the kind mentioned in the preceding paragraph.

General Background and Need for the Study

As stated previously, the past decade has been one of unrelenting change. There has also been an increased and widespread recognition that mathematics is tremendously important for the advancement of science. This recognition has been instrumental in prompting a re-examination not only of our mathematics curriculum but also of similar curricula in many countries of the world. Kemeny,[1] in a report to the

[1]John G. Kemeny, "Report to the International Congress of Mathematicians," The Mathematics Teacher, LVI (February, 1963), pp. 66-78.

International Congress of Mathematics at Stockholm, Sweden on August 15, 1962, emphasized that this critical examination of mathematics has been quite universal. Furthermore, this report states that:

> There are four areas of modern mathematics that are recommended by a majority of the reports. These are elementary set theory, an introduction to logic, some topics from modern algebra, and an introduction to probability and statistics.[2]

Much of the current discussion has been concerned with scholastic mathematics education. Perhaps, as Butler and Wren wrote in 1941:

> The reason why we hear so little of this criticism leveled at the teaching of calculus is simply because the matter of improving the teaching of college mathematics has as yet received almost no attention, at least in the way of published suggestions.[3]

However, one of the most serious problems confronting higher education today is the mushrooming college enrollments. These large enrollments have necessitated some harsh and critical scrutiny of large sized classes of undergraduates. Four approaches to this growing problem of dealing with large college classes are suggested by Kuuisto.[4] The approaches he proposes are mechanical teaching aids, the large lecture class, accelerated programs, and larger graduate schools. He includes television, teaching machines, and other audio-visual aids in the first of his proposals. He feels that these have been over-emphasized. He believes that too great an attempt has been made to

---

[2]Ibid., p. 69.

[3]Charles H. Butler and F. Lynwoods Wren, Teaching of Secondary Mathematics, (New York, 1941), p. 475.

[4]Allan A. Kuuisto, "What are the Most Effective Methods of Dealing with Larger Number of Students?", Higher Education in an Age of Revolutions, ed. G. Kerry Smith (Washington, 1962), pp. 173-176.

alleviate the college teacher shortage through the use of mechanical teaching aids. Furthermore, he feels that these teaching aids are best suited for adult education, "drill" courses, and off-campus services rather than the general college courses.

Of these four approaches proposed by Kuuisto, the only one, other than the mechanical devices, which should logically be considered as a method of teaching is the use of large lecture classes. The other two approaches constitute more of an administrative problem than a methodological problem. Kuuisto suggests that great care must be taken in selecting the professor for large lecture classes in order to ensure that the best possible teacher has been chosen.

There has been an increasing awareness in the last few years of the need for undergraduate college credit through examination and advanced placement. The consensus of opinion has been that if a college student has the desire and the ability, he should be given the opportunity to accelerate his college program, even though he may not have fulfilled the formal course requirements.

McKeachie[5] reports some studies which have indicated that those students who studied independently did better than those who did not. He is quick to point out, however, that this comparison was based on the use of tests on the textbook used for the particular course involved. These tests may or may not have tested all of the desirable outcomes of the course.

---

[5]W. J. McKeachie, "Procedures and Techniques of Teaching: A Survey of Experimental Studies," _The American College: A Psychological and Social Interpretation of the Higher Learning_, (New York, 1962), pp. 312-356.

The final approach to the higher educational problem of large enrollments as advocated by Kuuisto is a change or intensification in graduate education. The American graduate school is now in the process of expansion. It is rather early to determine whether or not this expansion has accomplished anything desirable in terms of the problem of large college enrollments. Nevertheless, it seems obvious that the nation's graduate schools are sufficiently committed to the doctorate as a basic qualification for college teaching. Much of this expansion may be in terms of a speeded-up doctoral program.

If Kuuisto's four approaches are accepted, then, as suggested previously, mechanical teaching aids and large lecture classes are the only two that are controllable by an individual college instructor or his particular department. The other two approaches are more dependent upon administrative decisions and, thus, are less likely to be manipulated by individual faculty members or small faculty groups. These considerations prompted the decision to explore the mechanical teaching aids and large lecture class approach.

The need for some feasible method of dealing with large numbers of students in higher education cannot be disputed. For example, the number of students in the Wisconsin State University system has more than doubled in the last four years.

> The system has grown more in the last four years than it did in the first 96, from 18,577 students to 38,592 this year. The growth rate is expected to be about 7,000 a year for the next two years.[6]

At Wisconsin State University at LaCrosse, the enrollment has increased

---

6 ________________, Wisconsin State Universities Report, XV (January, 1966), p. 3.

from a little more than eighteen-hundred in the fall of 1960 to over thirty-nine hundred in the fall of 1965. Similar increases have been common not only in the Wisconsin system but in institutions of higher education throughout the country. Thus, there appears to be a need to develop means of dealing with this tremendous influx of undergraduate students.

Of course, the growth in college enrollments would pose a less serious problem if there were not at the same time a shortage of capable teachers for these students. This is especially apparent in the discipline of mathematics. At Wisconsin State University at LaCrosse for example, the mathematics department would like to add at least two Doctors of Philosophy in mathematics and would add four if they were available. The department, however, will feel itself fortunate if it is able to secure one Doctor of Philosophy in mathematics.

Competence in undergraduate mathematics teaching has thus been equated with a doctorate in mathematics. While this may not be a completely fair assumption, it is nevertheless a criterion which is quite generally applied by educators throughout the United States. This criterion is applied not only by academic deans and other administrators but also by members of mathematics departments of American colleges and universities. Therefore, while the assumption that competence in undergraduate teaching is equivalent to a doctorate in mathematics may not be the fairest or most logical, it is an indication of the growing problem of staffing mathematics departments in American institutions of higher education.

The shortage of competent mathematics professors in American colleges and universities is inextricably combined with the large numbers of students prevalent in these colleges and universities today. Therefore, to attack the problem of large numbers of students is to attack the problem of such a shortage. Thus, any method for handling large numbers of students must be predicated upon the possible use of fewer, not greater, numbers of mathematics instructors. The assumption being made here, of course, is that higher educational institutions are interested only in competent mathematics professors.

Such a shortage and such large numbers of students in these colleges and universities are an outgrowth, or quite possible the cause, of America's concern for an improvement of teaching in its schools. Brearley suggests that "ultimately the improvement of teaching on the college level depends upon three groups of persons."[7] Brearley lists these three groups as persons as financial supporters, educational administrators, and the teachers themselves. This research was concerned with the teachers themselves.

Statement of the Problem

The purpose of this research was to evaluate a method of handling larger than normal sized classes in undergraduate mathematics with the aid of a type of mechanical teaching device. The particular mathematics course was a pre-calculus course entitled "Algebra and Trigonometry". This course is a prerequisite for the calculus sequence for those students with three years or less of mathematics preparation in high school. Most of the students involved in the research were freshmen, however, there were a few upperclassmen involved in this study.

---

[7] H. C. Brearley, "College Classroom Teaching: Problems and Procedures," Peabody Journal of Education, XXXVII (Summer, 1959), pp. 66-76.

Organization of the Study

During the 1965 fall semester at Wisconsin State University at LaCrosse, Wisconsin, three sections of Mathematics 109, Algebra and Trigonometry, were given eleven review sessions during the semester. One of the sections involved was taught by Professor X and the other two were taught by the writer. One of the sections taught by the writer was a large class that had a beginning enrollment of seventy-one students who completed the course. The writer's other section and Professor X's section had beginning enrollments of forty-three and thirty-eight students respectively. Thirty-six students completed the writer's smaller section and thirty-one students completed Professor X's section. Only data from those students who completed the course were used in this research.

The purpose of this investigation was to compare the increase in the level of achievement of the students in the experimental group who used the mechanical teaching device-review method with the increase in the level of achievement of the students in the control group who used the teacher-centered review method. While it is recognized by the writer that there are many outcomes that might have been properly considered as goals and objectives for a mathematics course of this type, the only outcome that was evaluated was subject matter achievement. Other related, concomitant, or aesthetic variables were not measured. Thus, no systematic effort was made to determine each students reaction to the review method to which the student was exposed. Also, no attempt was made to determine each students feelings as to his academic success because of the exposure to the particular review method.

The writer hopes that this study has produced some evidence which will contribute to a possible solution of the problems of a shortage of competent mathematics professors in American colleges and universities and of the large numbers of students in these colleges and universities. To accumulate this evidence effectively, two basic considerations were weighed before deciding upon a specific research design. The first of these considerations was the need for a valid research design. The second of these considerations was the decision to compare only the increase in achievement under the two different review methods. The need for results that would permit an evaluation of the increase in achievement under each of the two review methods dictated the use of some kind of standardized test. These tests were given to all of the students in these three sections during the second week of the semester. At the end of the semester different forms of the same standardized tests were given. These latter tests were given as the final examination for the course.

During the semester, each instructor taught his section at his own pace. No attempt was made to teach each section according to a rigid timetable. This was also true of the spacing of the various unit tests during the semester. The individual instructor was free to determine the placement of these unit tests.

At the start of the semester, each student in each of these three sections was assigned to either an experimental group or a control group. The experimental group consisted of those students who attended the synchronized tape recorder and slide projector review method. The control group consisted of those students who attended the teacher-centered, lecture and class-discussion review method. The placement in these groups was determined by the use of a table of random numbers. The students in each section were listed in alphabetical order with their last name first. Each student was then assigned a number on the basis of his position in this alphabetical order. Next, in a table of random numbers, a column of numbers was arbitrarily selected. Each section was assigned a different column from which the random numbers were selected. Since each section had less than one hundred students, only two digits were needed for each student. Since the table of random numbers consisted of numbers of more than two digits, each section arbitrarily assigned a set of two digits from each column. Then, using the appropriate column and the appropriate set for each section, one half of the numbers for each section were selected. If there was an odd number of students in a particular section, then one half plus one of the numbers for that section were selected. The students corresponding to the numbers so selected were assigned to the experimental group.

The procedure outlined perhaps was not completely ideal. However, as Lumsdaine states:

> Administrative conditions may require that the experimental instruction be performed in groups. In that event, random assignment of individuals to treatments may still be feasible. If so, this procedure may be greatly preferable to the block assignment of intact classroom groups because it reduces unwanted variability due to population variables not randomly distributed in the intact groups. This procedure is often much more feasible in two-group (two treatment) experiments than in cases where a larger number of treatments are employed, particularly if the experimental variation is introduced only for a single lesson or class period. Under these circumstances, students in a given classroom may be assigned randomly to two alternative groups, each of which is then instructed (and tested also, perhaps) in a separate room with one of the

alternative experimental instrument.[8]

Thus, the method of random selection of the experimental group indicated above appeared satisfactory.

An examination of the material to be covered during this pre-calculus course by Professor X and the writer determined the number of review sessions to be given. From this examination it was decided that eleven sessions were needed to review adequately the content of this pre-calculus course.

The three experimental groups met as a single group eleven times during the semester. Since these review sessions were in addition to the regular daily class session, the review sessions were held during the weekday evening hours. The review sessions were conducted on Tuesday or Thursday evenings from 7:00 to 8:00 P. M. The meetings were scheduled to precede the unit tests by a few days. At these meetings, the students reviewed a unit or a portion of a unit of course content using the mechanical teaching device. This device consisted of a tape recorder synchronized with a slide projector. The slides were automatically projected on the screen and were automatically changed by the tape on the tape recorder. The writer determined the length of time that the material on each slide was to appear on the screen. The material for the tapes and the slides was prepared by the writer. The voice on the tape was that of the writer.

The three control groups also met as a single group eleven times during the semester. These meetings were conducted at the same time as the meetings for the experimental groups. At these meetings of

---

[8] A. A. Lumsdaine, "Instruments and Media of Instruction," Handbook of Research on Teaching, ed. N. L. Gage (Chicago, 1963), p. 657.

the control group, the writer conducted a review session using the same classroom techniques that the writer used during his regular class sessions. Lecture techniques and formal and informal class discussion were used. The material covered in these sessions was the same as that covered in the experimental review sessions. The writer made a deliberate attempt to review exactly the same material as that covered by the mechanical device.

CHAPTER II

BACKGROUND FOR THE PRESENT STUDY

Introduction

The general method of teaching any class is usually based on the individual teacher's educational philosophy, the teaching objectives of the particular faculty involved, the objectives of the individual teacher, and the particular course being taught.

The classroom method most generally followed in the past has come to be called the teacher-centered classroom. This was characterized largely as a lesson-hearing, recitation method, where learning was considered a passive affair and teaching consisted mostly of telling, task fixing, and testing. As Dale states "We know the ingredients of training but perhaps not of education."[1] Dale goes on to say that:

> Educational material thrives on inference -- on what is not there. With training materials inference is at a minimum, the experiences to be undergone are all preplanned. You do not need to think, you accept and imitate.[2]

If we accept this distinction between education and training, it seems quite clear that the old teacher-centered classroom stressed a method which placed major emphasis on training, and not on education.

---

[1]Edgar Dale, "New Techniques of Teaching," The Two Ends of the Log, ed. Russell M. Cooper (Minneapolis, 1958), p. 193.

[2]Ibid.

As teachers became more aware of discoveries and advances in educational psychology and of a broader concept of citizenship which placed the emphasis upon group interaction, there came a new method of teaching called the pupil-centered classroom. Several new techniques have emerged from this pupil-centered approach to teaching. Two of the better known of these techniques are the project technique and the problem solving technique. Both of these have had significance in influencing mathematics teaching methods.

The present world is considerably different from that of our fathers' and grandfathers'. As Trippet sees this world of today:

> It is at once a larger and a smaller world. It spins more rapidly, is more densely populated, more interdependent and interrelated. It encompasses quasi-miraculous sources of power. It engenders a host of tensions and conflicts and seemingly insoluble human problems.[3]

The means of dealing intelligently with the present and future problems of today's world must be one of the major concerns of education. Rosenbloom voices this thought when he says that:

> I take as the main goal of education that of preparing the student to take his place in the adult world. To do this he must understand the world around him -- both the world of nature and the society in which he lives -- and must discover his own abilities and interests, he must develop these abilities as far as possible, and he must acquire a scale of values which inspires him to make the best possible use of his talents. Furthermore, he must be prepared as a future parent to transmit, as his children's first and most influential teacher, the cultural heritage of his society.[4]

According to Price, who states this thought more fully:

---

[3]Byron K. Trippet, "Are Fundamental Changes Required in Higher Education?", <u>Goals for Higher Education in a Decade of Decision</u>, ed. G. Kerry Smith (Washington, 1962), p. 27.

[4]Paul C. Rosenbloom, "The Role of Mathematics and Science in a General Education," mimeographed article, (Minneapolis, 1959), p. 1.

> In the future, a college education must be considered general education for the great bulk of the population, whereas in the past a college education was designed for the privileged few. This fact will have an important bearing on the mathematics that is taught in our undergraduate college and university mathematics courses in the future. The mathematics courses that will be taught will be designed to help members of the general public to make a living and to discharge their duties as citizens. In order to provide the general education that will be required, colleges and universities must make certain that their students reach a higher level in mathematics than formerly.[5]

Thus, the goals for education stated previously by Rosenbloom must, following Price's reasoning, also be Rosenbloom's goals for higher education.

In American education today there appears to be more concern with an acceleration of the education of some students. In higher education, advanced placement is becoming more and more of a routine procedure. Pressey[6] argues that since the most outstanding creative work is done by quite young people, for them to remain in schools when they could have finished formal education earlier is a waste of their precious creative talents. He also argues that by encouraging young people to accelerate their education and to move more quickly into their professions, they might be more inclined to defer marriage until they were actually at work. His argument is given further emphasis by the careers of distinguished men such as Nobel Prize winners who finished their degrees early. Pressey[7] further states that there is little evidence to support enrichment programs as a means of meeting the needs of

---

[5] G. Baley Price, "New Perspectives on Teaching Mathematics," Higher Education in an Age of Revolutions, ed. G. Kerry Smith (Washington, 1962), pp. 100-101.

[6] Sidney L. Pressey, "Educational Acceleration: Occasional Procedure or Major Issue," The Personnel and Guidance Journal, XV (September, 1962), pp. 12-16.

[7] Ibid.

bright students. He also states that there is considerable data which suggest that most students in the upper twenty per cent of the ability range could profitable speed up their education with no attendant social maladjustment.

There appears then to be sufficient evidence to support the conclusion that higher education should encourage students to concern themselves with ways in which they might profitable accelerate their education. That is, students in higher education should be encouraged to develop attitudes which will enable them to pursue their educational goals with a greater amount of initiative and independence.

In February of 1954 the convention of the American Educational Research Association conducted a round table in Research in Science and Mathematics. Mallinson[8] identified the five most needed research investigations in the teaching of mathematics, as determined by this round table, to be:

> (a) The identification of the concepts and functional competence in mathematics needed for the general education of all students at the secondary level.

Mallinson,[9] the chairman of the round table, reported that most of the participants were of the opinion that many of the studies in the area of objectives for the teaching of mathematics deal with mathematical skills such as the ability to compute and to do square root rather than with concepts, competencies, and understandings. Further, the members of the round table felt that major emphasis is

---

[8] C. G. Mallinson, "The Five Most Needed Research Investigations in the Teaching of Science and Mathematics," School Science and Mathematics, LIV (June, 1954), pp. 428-430.

[9] Ibid.

needed on identifying and defining those concepts, competencies, and understandings that may be suitable as objectives for general education.

(b) The background in mathematics needed for teaching courses in mathematics for general education.

In Mallinson's[10] report, he noted that recent surveys indicate that few if any colleges offer courses designed to aid teachers in teaching such courses in mathematics for general education.

(c) The development of tests that present situations in which the methodology of mathematics is tested rather than the skills of mathematics.

As might be inferred from the preceding discussion, Mallinson[11] reported that the round table members' definition of "the methodology of mathematics' was meant to be the concepts, competencies, and understandings of mathematics which were previously considered. The round table, of course, felt that these were desirable results of the teaching of mathematics, and that if these were considered to be results by the majority of mathematical educators, then tests should be designed to measure them.

(d) The development of techniques for teaching mathematics inductively and for teaching students to think mathematically.

The participants, Mallinson[12] reported, agreed that while it is not difficult to state the objectives that are desirable in these two areas, particular methods of implementing these are lacking.

(e) The need for non-technical publications that will summarize the implications and practical applications of research in the teaching of mathematics.

---

[10]Ibid., PP. 428-430

[11]Ibid.

[12]Ibid.

Mallinson[13] reported that the consensus of opinion of the round table members was that most classroom teachers, and some professional educators, have little opportunity to analyze the more or less technical literature that contains the research investigations in science and mathematics. Thus, the round table strongly urged that the American Educational Research Association prepare "laymen's reviews" in these two areas, such as the one recently prepared for the field of teaching.

Review of Relevant Literature

In the past decade, there has been little completed research that fits nicely into any of these five categories. Burkhard,[14] in a thesis which was part of the requirements for a Doctor of Philosophy degree from Columbia University, published "A Study of Concept Learning in Differential Calculus" in which she sought to determine the methods and materials needed to increase the students understandings of concepts in calculus. In her study, the mathematical literature was searched to determine the nature and important aspects of the concepts of differential calculus. Two hundred thirty-five students comprising nine differential calculus classes were involved in the experimental portion of this study. The students in the experimental group were taught with a greater emphasis being placed upon understanding the concepts of the calculus. Two classes were taught in the conventional manner with the major emphasis on skills and problem solving. The differences observed

---

[13] Ibid., pp. 428-430

[14] Sarah Burkhard, "A Study of Concept Learning in Differential Calculus," Dissertation Abstracts, XVI, No. 2 (Columbia University, 1956).

between the two groups of students were primarily in the area of quality of concept, with the greater grasp of the concepts coming from the experimental group. There were, however, no statistically significant results obtained from this study.

A study conducted by Smith[15] is in this same general area of mathematical concepts. In this study an attempt was made to compare an algebraic and a geometric method of teaching a college general mathematics course. A total of one hundred forty-one students were divided into three groups. The same instructor taught each group, using the algebraic method for the first group, the geometric method for the second group, and the third group served as a control class. Although the results did not uniformly favor either method of presentation, the advantages of the geometric method outweighed those of the algebraic method for most students.

In the last several years there has been a great deal of discussion concerning the relative merits of acceleration and enrichment. Hyman, writing in the Journal of Higher Education, proposes:

> That the Fund for the Advancement of Education conduct an "experiment" in which a thousand or so students (would) receive syllabuses, textbook, review books, and library cards, while an equal number of students (would) attend classes.[16]

Perhaps this would be an impractical plan, but Williamson[17] has done

---

[15]Roland Frederick Smith, "An Experimental Comparison of Two Liberal Arts Courses in General Mathematics at Syracuse University," Dissertation Abstracts, XV, No. 6 (Syracuse University, 1955).

[16]Lawrence W. Hyman, "Advancing Education by Eliminating Classes," Journal of Higher Education, XXXII (April, 1961), pp. 213-215.

[17]Robert Gordon Williamson, "A Theory of Learning and Its Application to a Class in College Mathematics," Dissertation Abstracts, XVI, (University of Maryland, 1956), p. 395.

some research which seems to indicate that Hyman's idea is not altogether impractical.

Williamson[18] completed a dissertation on "A Theory of Learning and Its Application to a Class in College Mathematics". In this study he attempted to use a philosophical approach to deduce an original method for teaching college subject matter. He reviewed previous learning theories and developments in modern science and mathematics. From this review he developed a theory of learning. A procedure was devised that applied this theory as a highly individualized method of instruction with particular emphasis on student self-involvement and student-teacher communication. A single class in freshman college mathematics was taught using this procedure and measuring results in relation to certain selected factors. This method permitted students to proceed at their own rate under individualized instruction. The study showed that student ratings and evaluations indicated a preference for this method over the conventional method of teaching and that significant gains in subject matter skill were achieved. It is worth noting, however, that the results of this study did not indicate whether this gain was in relation to the student's subject matter skill at the beginning of the course or in relation to what might be expected if they had been taught in the conventional manner. Also, the use of the words "subject matter skills" was not defined, and unless the words "subject matter skills" denote those concepts, competencies, and understandings which the members of the round table felt were the highly desirable results of college general mathematics classes, then

[18] Ibid, p. 395.

Mallinson and the other members of the round table would agree that the study is not particularly relevant to the question of how efficiently to secure these highly desirable results.

The exploding college population has been a source of much concern for all educators. This problem has been attacked by the members of Cornell University's mathematics department who sought a solution to the problem of teaching freshmen with vastly different interests and backgrounds in mathematics. In a study reported in School and Society,[19] the members of Cornell University's mathematics department, at the beginning of the school year, divided their first-year calculus students into three groups on the basis of their mathematical and verbal aptitude test scores. If any obvious or necessary shifts were needed, these were made after three weeks of the session. According to the chairman of the department of mathematics, the new program allowed better equipped students to move faster than before and students with less preparation to get more help. At the same time no lowering of academic standards was permitted. The approximately one thousand students who registered for first-year calculus were placed in fifty class sections of about twenty students each. Ten of these sections received a course that covered more material and theory than the former first-year course. After the first term, twenty of these students were then transferred to a special section that covered even more ground. The other forty sections received the typical calculus course meeting three hours a week, but the "lower" ten of these sections received an extra hour of class work a week without college

[19]"Individual Differences in College Mathematics," School and Society, LXXXIV (December 8, 1956), p. 204.

credit. The students in these ten sections seemed pleased with the extra instruction, particularly since many students had formerly hired tutors for extra work. However, there was no significant statistical change in achievement under this system.

Almost all colleges have at some time or other had some form of tutoring service. Hampton Institute has conducted a rather unique tutorial service. Hawkins[20] writes about "A Volunteer Tutorial System" in the Phi Delta Kappan. In this undertaking at Hampton Institute in Virginia, the tutors were required to have at least a general average of B. The tutors were also required to have at least a B average in the subject to be tutored and a willingness to render service without receiving financial compensation. In addition to these requirements, a recommendation from the chairman of the department of the subject being tutored was required. It was felt that this arrangement gave educational guidance, developed student leadership, and provided an opportunity for the gifted students to utilize their talents in the service of others.

There have been many words written in recent years about the use, or possible use, of television as a teaching tool for relieving the shortage of classrooms and teachers. Many problems have arisen in the use of the medium, and many questions have been posed concerning the results of the use of this device. Perhaps one of the major worries is unconsciously implied by the following quotation from Dale:

> One hazard of mass education, of the use of larger and larger classes, is that the instructor must exactly define the right

[20] Thomas E. Hawkins, "A Volunteer Tutorial System," Phi Delta Kappan, XL (January, 1959), pp. 168-169.

> answers and give grades and marks on this basis. The student answers questions, but he does not question answers. You may say that this is not a necessary concomitant of huge classes, but the mere fact of size makes it difficult to do anything else. We shall not get differentiated responses if differentiated responses are not rewarded by those who make up the tests.[21]

Huge classes are not a problem according to Carpenter.[22] In reporting on the results and impressions of the use of television at Pennsylvania State University, he states that the research people there who were connected with the use of television in teaching felt that:

> The more strictly an experiment is controlled, the greater the probabilities that there will be non-significant statistical differences between scores of students taught by television and those taught conventionally.[23]

Furthermore, these research people felt that:

> We have the means in television for making a substantial contribution to the solution of the "quantity" problem in American education. It remains to be shown "how and to what extent" teaching by television can contribute to the related problem of improving the "quality" of college and university teaching.[24]

Benner and Rogers[25] report on a television experiment at the University of Houston in the May, 1960 issue of The Mathematics Teacher. In this study, plane trigonometry was taught to approximately two hundred and fifty students. The basic features of the plan were:

---

[21]Edgar Dale, "New Techniques of Teaching," The Two Ends of the Log, ed. Russell M. Cooper (Minneapolis, 1958), p. 193.

[22]C. R. Carpenter, "Teaching and Learning by Television," The Two Ends of the Log, ed. Russell M. Cooper (Minneapolis, 1958), p. 217.

[23]Ibid.

[24]Ibid.

[25]Charles Benner and Curtis A. Rogers, "A New Plan for Instructing Large Classes in Mathematics by Television and Films," The Mathematics Teacher, LIII (May, 1960), pp. 371-375.

(a) There were six members of the mathematics department who gave twenty-seven lectures which were checked by members of the mathematics department and which were forty-four minutes in length;

(b) two lectures were given each week, with each:
film-shown twice over open circuit television (morning and evening) and twice by projectors in viewing rooms of the Audio-Visual Center (afternoons and Saturday mornings) - provided on campus for all televised lectures.

All students enrolled for credit were supplied with a television supplement which contained routine instructions for the course, a list of study aids, additional explanation, and an incomplete set of notes on all lectures.[26]

(c) the students completed these notes as they viewed the lecture;

(d) the students were divided into sections of approximately thirty students which were required to meet one hour per week in a conference session with a member of the mathematics department where the lectures were discussed, questions answered, homework collected, and reviewing for examinations done.

(e) In addition to these conference sessions, ten hours of help sessions were scheduled, at which a student assistant was in charge to answer questions about specific problems;

(f) two comprehensive examinations of two hours length were preceded by a live television review and two, more limited, examinations which were given in the weekly conference session.

The University of Houston mathematics faculty devised a plan for producing the television lectures which consisted of:

(a) the assigning of a topic to a lecturer who then produced rough notes which were presented to the other members of the mathematics faculty for criticism;

(b) the refining of these rough notes, using the criticisms given, into production form;

(c) the taping of the actual lecture which was then made available to the mathematics department, along with the production notes, for further criticism;

---

[26]Ibid, pp. 371-375.

(d) the producing of the final production notes using the two sets of criticisms already given.

In addition to problems from the production end of this television venture, certain student difficulties also became apparent. The taking of notes became more restricted, since it was difficult to catch up if the student once fell behind as there was less material in view at any one time. This difficulty was alleviated somewhat by the distribution of the incomplete lecture notes. It was also impossible for the student to interrupt and ask a question of the lecturer. This was solved in part by the lecturer himself in anticipating the usual questions when preparing his script and also by suggesting to the students that they write questions down and bring them up in the conference sessions. Perhaps the most immediate difficulty was the students' lack of experience with this form of instruction. Most of this course's requirements were completely voluntary and the necessary self-discipline was often lacking, at least in the beginning. Orientation procedures, such as a preliminary telecast and the distribution of an orientation pamphlet, seemed to reduce this problem somewhat. The members of the mathematics department felt that this problem would continue to decrease in scope as more and more students became familiar with this instructional technique.

Another study, more limited in scope than the two previously reported, "was conducted during the winter quarter, 1962, at George Peabody College for Teachers, Nashville, Tennessee".[27] Three different

[27] Horace E. Williams, "A Study of the Effectiveness of Classroom Teaching Techniques Following a Closed-Circuit Television Presentation in Mathematics," The Mathematics Teacher, LIV (February, 1963), p. 94.

classes of a general mathematics course received a twenty-five minute presentation over closed-circuit television. The same presentation was given to the three classes at the same time, with three different post-television methods practiced in the classrooms.

In the first method, the television lecturer remained on television and answered questions relayed to him by an instructor in the classroom. The television presentation was made from a studio with all three classrooms some distance removed from the television studio. This necessitated the need for the relay to the television lecturer. During this post lecture presentation, the stress was on informal and personalized instruction with no new material presented.

For the second method the classroom instructor simply answered questions asked by the students after the television lecture was presented. The instructor attempted to clear up any areas of difficulty revealed by the students' questions.

In the third method, the classroom instructor used the rest of the period following the television preparation to approach the lecture topics from a different point of view. He attempted to accomplish this not only by using lectures, but also by using illustrations and questions which he asked the students.

After evaluating the data derived from this experiment, the following conclusions were drawn:

1. There was no apparent difference in the effectiveness or of the retention by either of the three methods;

2. Since there were very small differences in these three methods any one may be safely used by any one using a television presentation.

McKeachie[28] gives an excellent report on the procedures and techniques of teaching in a chapter of The American College. He stresses the difficulty of comparing different teaching methods. He asserts that this difficulty is due mainly to ineffective evaluation, changes in student motivation, and different levels of teacher effectiveness in terms of particular course objectives. Consequently, it is difficult to compare positively and carefully two different methods of instruction. A general comparison then of the "automated" (that is, television, films, and programmed materials) teaching techniques and teacher procedures reveals no clear cut advantage for either of these "processes". For differing objectives, student characteristics and materials, each of the techniques indicated above has been shown to be superior for a particular situation. Much remains to be done, and the unifying thread running through all of the literature is the need for more and better articulated experimentation.

While there has been much publicity about the need for the revision of the secondary mathematics curriculum, there has been very little attendant publicity about the collegiate mathematics curriculum. The layman at least, has been unaware of any need for a revision of collegiate mathematics, let alone any notice that any change is being made.

A comparison of almost any college or university catalog of today with the same institution's catalog of ten or fifteen years ago will convince the examiner that change has taken place. The traditional

[28] W. J. McKeachie, "Procedures and Techniques of Teaching: A Survey of Experimental Studies," The American College: A Psychological and Social Interpretation of The Higher Learning, (New York, 1962), pp. 312-356.

college algebra, trigonometry, plane analytic geometry, solid analytic geometry, differential calculus, and integral calculus sequence of ten or fifteen years ago has often been replaced by a sequence generally consisting of elementary analysis (algebra and trigonometry), an introduction to calculus (including some advanced topics in algebra as well as plane analytic geometry), differential calculus, and integral calculus (including solid analytical geometry).

It is difficult to determine whether this change in the collegiate mathematics curriculum preceded, or was preceded by, the corresponding change in the secondary and elementary mathematics curriculum. Regardless of the order of change, this change is taking place. Furthermore, this collegiate change is a dynamic process. This is quite significant and indicates that the collegiate educator, as well as the secondary educator, now recognizes that the need for curriculum revision is ever present.

This discussion has been concerned with the college mathematics curriculum. We have however, concerned ourselves only with the needs of the prospective teachers of mathematics. But are the objectives for the education of teachers of mathematics the same as the objectives for other students with an interest in mathematics? Dubisch, of the University of Washington, puts the students of mathematics in four classes which are as follows:

1. Those who plan to take an advances degree in mathematics.
2. Those who plan to teach elementary or high school mathematics.
3. The prospective majors in subjects that use mathematics extensively, such as, engineering, physics, and chemistry. (We may also include in this group those mathematics majors in college who plan to enter an applied field such as computers after their baccalaureate degree).

4. Those students who are taking mathematics as an elective either because of inherent interest in the subject (without planning to major in it or to apply it) or because of a college requirement for entrance or graduation.[29]

While Dubisch recognizes that each of these four groups has different special needs, he feels that there is enough similarity in their requirements to have them placed in the same class in college. This is especially true, since he advocates the teaching of mathematics with the emphasis on the thinking process. Thus, Dubisch seems to be saying that for each of these groups, the dominant factor in teaching each group should be the nature of mathematical structure. This need for an emphasis on structure has been advocated by Bruner[30] in the Process of Education and Mayer[31] in The Schools.

There appears then to be an enigma in the literature concerning undergraduate mathematics. The need for extensive, quality preparation for those who are going to teach or to use mathematics in their post collegiate careers cannot be denied. Contrast this with the shortage of qualified collegiate mathematics professors. Woven into the problem is a need to continually strive for better and better college instruction.

Brearley[32] lists six background factors for improvement of college instruction. They are:

---

[29] Roy Dubisch, "The Aims of Teaching Mathematics," The Teaching of Mathematics, (New York, 1963), p. 7.

[30] Jerome S. Bruner, The Process of Education (Cambridge, 1961), p. 11.

[31] Martin Mayer, The Schools (New York, 1961), p. 234-266.

[32] H. C. Brearley, "College Classroom Teaching: Problems and Procedures," Peabody Journal of Education, XXXVII (Summer, 1959), pp. 66-76.

1. policies and emotional and intellectual climate;

2. type of classroom;

3. previous training and intellectual maturity of students;

4. scholarship, personality, and relative skills of teacher;

5. nature of material and educational purpose of instruction;

6. aim or direction of teaching;

    a. information,
    b. skills,
    c. insight or understanding,
    d. attitudes or points of view.

Using these as criteria for improving instruction, the literature seems to suggest that the teaching of a pre-calculus college mathematics course to large classes using some type of mechanical teaching device is feasible and that this is possible while improving the mathematical instruction involved.

This conclusion must be qualified by the following admonition of Brown and Thornton:

> The use of teaching machines or programmed book materials in college instruction presents special problems which merit consideration by the instructor.
>
> If programmed materials are to be used, programs will need to be found (not easy, but becoming easier) which meet one's expectations and specifications.
>
> Programs requiring nonportable machines (as opposed to book types or individual, low-cost machines which can be carried about by the user) will require special rooms; someone must be responsible, too, for loading machines, keeping them functioning, and reclaiming and analyzing residual answer sheets. Programmed book materials, now becoming more common, have the advantages of portability, flexibility, low cost, and individualization.
>
> The instructor who assigns programmed materials to students must give special consideration to changes this action is likely to require in the usual patterns of in-class and out-of-class activities. Instructors who use programmed materials for example, sometimes find that it is possible to devote more

> time in class to discussion and explanation of confusions or misunderstandings arising from out of class study of programmed materials and less to lecturing or other instructor-presentation techniques.
>
> Instruction must be given students in how to use programmed materials and equipment.[33]

One must recognize, however, that the press of college enrollments may be a blessing in disguise. It has forced the higher education community to examine different methods of handling these large numbers. But if, as Baskin suggests:

> Our job is to shift the focus in the college classroom so that the student begins to look more and more to his own resources for his learning.[34]

then these examinations may improve the whole level of collegiate mathematics instruction.

---

[33]James W. Brown and James W. Thornton, Jr., College Teaching: perspectives and guidelines, (New York, 1963), p. 188.

[34]Samuel Baskin, "Independent Study: Methods, Programs, and Whom?" Higher Education in An Age of Revolutions, ed. G. Kerry Smith (Washington, 1962), p. 65.

CHAPTER III

DESIGN OF THE PRESENT STUDY

Introduction

The basic objective of this research was an attempt to discover whether there was any significant difference between a review method featuring a mechanical teaching device and a review method based upon a teacher-centered lecture-discussion group. To achieve this objective it was deemed essential to consider the following questions:

1. What are the different types of experimental designs which are applicable to this research?

2. Is there an experimental design which is better suited to this research than others?

3. Are there statistics which will better enable one to make an intelligent decision concerning the difference or lack of difference between the two methods?

This chapter gives some of the more essential characteristics of a good experimental design. These characteristics were then used to evaluate several different experimental designs. This evaluation was the criterion used to determine the design of this research. After the experimental design had been selected, an examination of the different statistical treatments was conducted. This examination attempted to ensure the use of statistical evaluations that would allow the most intelligent decision to be made concerning the relative merits of the two review methods.

Theoretical Background

It is difficult to make an assertion concerning the purpose of all educational experiments. One can only make a supposition based upon the available evidence, and then await the scholarly suggestions of his colleagues and associates.

Lindquist states that:

> The major purpose of psychological experiments is to describe the effect of certain experimental "treatments" upon some characteristic of a particular population, or to test some hypothesis about this effect.[1]

His use of the term "treatment" refers to any variation in procedures which are to be observed and evaluated. While Lindquist's statement explicitly says "psychological experiments", the context of the remarks in which this statement was made makes it clear that he implicitly included educational experiments in this statement. Thus, his statement will be accepted as the purpose of educational experiments.

In general, experimental results will vary from subject to subject, experiment to experiment, and treatment to treatment. The results are influenced by variations in many different factors. Thus, the observation from a single experiment must be regarded as simply an estimate of the actual effect of the experiment. The actual effect is the result which would have been obtained if the experiment had been perfectly controlled and if it had involved all the members of the population being studied. Thus, as Lindquist states,:

---

[1]E. F. Lindquist, "Design and Analysis of Experiments in Psychology and Education," (Cambridge, Mass., 1956), p. 1.

> The usefulness or value of the experiment, therefore, depends upon two major characteristics of the estimate obtained:
>
> 1. its freedom from bias, and
>
> 2. its precision.
>
> An estimate may be said to be free from bias to the degree that its <u>average</u> value for an increasing number of similar experiments tends to approach the "true" value. The precision of the estimate depends upon the variability of such estimates for such a series of experiments - the less variable the estimates, the more precise is any single estimate.[2]

Most educational research will be concerned with an attempt to develop a more dynamic and viable theory of education and learning. Thus, educational experiments should have as their main objectives the descriptions of the effects of the treatments <u>and</u> the testing of specific hypotheses concerning the true effects of the treatments. Generally, the simplest possible hypothesis which will explain the observations is tested first. This hypothesis is usually that there is no true difference between the experimental treatments. Thus, the purpose of most experiments is to test a "null" hypothesis. Other hypotheses will be considered only if the "null" hypothesis is rejected.

Even if the hypothesis to be tested is true, the experimental observations cannot be expected to agree completely with the hypothetical observations. Lindquist states:

> Noting the discrepancy between the observed effect and the hypothetical true effect, we ask, is this discrepancy too large to be reasonably attributed to "error", - too large to enable us to retain the hypothesis? If so, just how confident may we be that the hypothesis is false? If the experiment has been properly designed, we can supply objective and quantitative answers to these questions. Thus <u>a major objective of the design of an experiment is to make such answers possible</u>.[3]

---

[2] Ibid., p. 2.

[3] Ibid., p. 6.

Using these criteria then, Lindquist asserts that a good experimental design must:

1. --insure that the observed treatment effects are unbiased estimates of the true effects.

2. --permit a quantitative description of the precision of the observed treatment effects regarded as estimates of the "true" effects.

3. --insure that the observed treatment effects will have whatever degree of precision is required by the broader purposes of the experiment.

3. --make possible an objective test of a specific hypothesis concerning the true effects; that is, it will permit the computation of the relative frequency with which the observed discrepancy between observation and hypothesis would be exceeded if the hypothesis were true.

5. --be <u>efficient</u>; that is, it will satisfy these requirements at the minimum "cost", broadly conceived.[4]

In considering various experimental designs, these criteria will be used to determine the desirability of the different designs.

Several basic designs have been listed by Lindquist.[5] These are simple-randomized designs, treatments by levels designs, treatments by subject designs, random replications designs, factorial designs, and groups-within-treatments designs.

The simple-randomized design is one of the most important experimental designs. It is not only used by itself, but it is also used in many of the more complex designs used for experimental research. In this design each treatment is independently administered to a different group, with each group drawn at random from the same parent population.

---

[4]Ibid., p. 6.

[5]Ibid.

In the treatments by levels design, the treatments are administered to samples that have been "paired" with respect to a particular "control" variable. This design increases the precision of the treatment comparisons by the use of this "matching up" process. The null hypothesis usually tested is that the population mean is the same for all treatments.

The treatments by subjects design has the treatments administered in succession to the same subjects, and not to different groups of subjects. The use of this design increases the precision of the experiment through the elimination of between-subject differences which are a source of error. This design can rarely be used in learning experiments since the experimenter must be interested in the cumulative effects of the treatments.

The random replications design is generally used when the population consists of a finite number of groups of which only a few may be represented in any one experiment. The experiment is independently duplicated for each of the groups. The design employed in each experiment may be the simple-randomized design or some other design. From a tests of significance standpoint, the random replications design is essentially the same as the treatments by subjects design.

The factorial design allows one to study several experimental variables simultaneously. This increases the precision of the experiment and permits an examination of the possible interaction between treatments and levels. There is a great similarity between the factorial design and the treatment by levels design. If a variable is introduced and it is not known in advance, if the second variable is related to the first, then the design to be used will be a factorial one.

The groups-within-treatments design is used when the purpose of an experiment is to generalize for a population which consists of many subpopulations and it is not possible to duplicate the experiment for each of the subpopulations. If the groups are not of the same size, then it is usually desirable to give all the groups the same weight in the treatment comparisons even though the groups differ in size.

The designs listed above are rarely used in exactly the same form as described in any actual research. Most research designs employed in actual practice are combinations of the basic designs presented previously.

The selection of a particular design must also be concerned with the question of validity. Campbell and Stanley make a distinction between internal and external validity. They state that internal validity must answer the question "Did in fact the experimental treatments make a difference in this specific experimental instance?"[6] External validity, they say, must concern itself with the question "To what populations, settings, treatment variables, and measurement variables can this effect be generalized?"[7]

Campbell and Stanley then list twelve different classes of variables which must be controlled in the design of the experiment. Failure to control these variables might produce effects which may influence and interact with the experimental stimulus. Their list of variables is as follows:

---

[6]Donald T. Campbell and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," <u>Handbook of Research on Teaching</u>, ed. N. L. Gage, (Chicago, 1963), p. 175.

[7]Ibid.

1. History, the specific events occurring between the first and second measurement in addition to the experimental variable.

2. Maturation, processes within the respondents operating as a function of the passage of time per se (not specific to the particular events), including growing older, growing hungrier, growing more tired, and the like.

3. Testing, the effects of taking a test upon the scores of a second testing.

4. Instrumentation, in which changes in the calibration of a measuring instrument or changes in the observors or scorers used may produce changes in the obtained measurements.

5. Statistical regression, operating where groups have been selected on the basis of their extreme scores.

6. Biases resulting in differential selection of respondents for the comparison groups.

7. Experimental mortality, or differential loss of respondents from the comparison groups.

8. Selection-maturation, interaction, and so forth, which, in certain of the multiple-group quasi-experimental designs, might be mistaken for the effect of the experimental variable.

9. The reactive or interaction effect of testing, in which a pretest might increase or decrease the respondent's sensitivity or responsiveness to the experimental variable and thus make the results obtained for a pretested population unrepresentative of the effects of the experimental variable for the unpretested universe from which the experimental respondents were selected.

10. The interaction effects of selection biases and the experimental variable.

11. Reactive effects of experimental arrangements, which would preclude generalization about the effect of the experimental variable upon persons being exposed to it in nonexperimental settings.

12. Multiple-treatment interference, likely to occur whenever multiple treatments are applied to the same respondents, because the effects of prior treatments are not usually erasable.[8]

---

[8]Ibid., pp. 175-176.

Of these twelve variables, the first eight pertain to internal validity and the last four to external validity.

Campbell and Stanley then discuss sixteen different designs which they categorize as pre-experimental, true experimental, and quasi-experimental designs. Only the three experimental designs which they classify as true experimental designs will be discussed here since they appear to be the ones which have the most to recommend them for research of the type being conducted. An appraisal of the sources of invalidity for these designs is given in TABLE I. This is a portion of a similar table prepared by Campbell and Stanley. Particular emphasis should probably be placed upon the footnote to this table. The table is only meant to be a guide for the reader, and not a "hard and fast" rule which must be accepted as "truth".

In discussing these three true experimental designs:

> An X will represent the exposure of a group to an experimental variable or event, the effects of which are to be measured; O will refer to some process of observation or measurement; the X's and O's in a given row are applied to the same specific persons. The left-to-right dimension indicates the temporal order, and X's and O's vertical to one another are simultaneous.[9]

The symbol R will indicate a random assignment to the different treatment groups. Of the three true experimental designs listed by Campbell and Stanley,[10] the Pretest-Posttest Control Group Design is the more widely used. The form of this design is as follows:

$$\begin{array}{cccc} R & O_1 & X & O_2 \\ R & O_3 & & O_4 \end{array}$$

[9]Ibid., p. 176.

[10]Ibid.

This design controls for all sources of internal validity. It does not, however, control for any of the sources of external validity. The most serious deficiency here is that it does not control for the interaction between testing and the experimental variable. While this design does not control for interaction between selection and the experimental variable, this is not too important in research on teaching since the population to be studied is a captive one. Generalization to the average citizen is not necessary. This design does not control for reactive arrangements. This phenomenon discourages generalization when the experiment is conducted in a setting which is patently artificial. The solution to this problem is to disguise the experiment as much as possible. This is not as difficult in research on teaching as it is in other forms of psychological experimentation.

The Solomon Four-Group Design has the following form:

R	O_1	X	O_2
R	O_3		O_4
R		X	O_5
R			O_6

This design has all of the controls of the Pretest-Posttest Control Group Design and, in addition, it controls for the interaction between testing and experimentation. Thus, generalizability is increased. By comparing O_2 with O_1, O_2 with O_4, O_5 with O_6, and O_5 with O_3, the effects of experimentation can be ascertained more completely. This design does not control for reactive arrangements and interaction of selection and experimentation. However, the discussion concerning the Pretest-Posttest Control Group Design is as pertinent for this design.

TABLE I

SOURCES OF INVALIDITY FOR TRUE EXPERIMENTAL DESIGNS

	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple - X Interference
True Experimental Designs: Pretest-Posttest Control Group Design R O X O R O O	+	+	+	+	+	+	+	+	-	?	?	
Solomon Four-Group Design R O X O R O O R X O R O	+	+	+	+	+	+	+	+	+	?	?	
Posttest-Only Control Group Design R X O R O	+	+	+	+	+	+	+	+	+	?	?	

Note: In the tables, a minus indicates a definite weakness, a plus indicates that the factor is controlled, a question mark indicates a possible source of concern, and a blank indicates that the factor is not relevant.

It is with extreme reluctance that these summary tables are presented because they are apt to be "too helpful," and to be depended upon in place of the more complex and qualified presentation in the text. No + or - indicator should be respected unless the reader comprehends why it is placed there. In particular, it is against the spirit of this presentation to create uncomprehended fears of, or confidence in, specific designs.[11]

[11] Ibid., p. 178.

The Posttest-Only Control Group Design has the form

$$\begin{array}{lll} R & X & O_1 \\ R & & O_2 . \end{array}$$

This design also controls for all sources of internal validity and for the interaction of testing and experimentation. It does not measure the effect of interaction of testing and experimentation, however. This is often not a problem since one is often only interested in answering the question of whether or not there is interaction and not of how much interaction. Since there is no pretest in this design, it seems logical to assert that there is not as much reactive interference in this design as in the others. One cannot say, however, that there is no reactive arrangements, only that there appears to be less in this design than in the other two.

Hypotheses

The hypotheses of this study concern the achievement of the experimental group, using the mechanical device for review, versus the control group, using the teacher-oriented type of review method. These hypotheses will deal with the lack of significance between the two levels of achievement. They will be stated in terms of means and variances of the different groups under consideration.

The hypotheses to be tested in this research are as follows:

1. There is no difference between the means of the experimental and control groups.

2. There is no difference between the variances of the experimental and control groups.

These hypotheses will be tested under the assumption that:

1. Each treatment group will be randomly selected from the appropriate subpopulation of the population tested.

2. The distribution of these subpopulations will be normal.

3. All of these distributions will have the same variance (σ^2).

To further illustrate these assumptions, suppose that we wished to test the equality of six population means using six independent random samples. Thus we wish to test the null hypothesis

$$H_0: \mu_1 = \mu_2 = - - - - - = \mu_6$$

against the alternate hypothesis

$$H: \text{at least two means are not equal}$$

where μ_j, $j = 1, 2, - - -, 6$ is the mean of the j^{th} population. The j's represent the six different treatments given to the six different populations. The size of the population that has received treatment j will be denoted by n_j and x_{ij} will denote the i^{th} observation receiving treatment j, where $i = 1, 2, - - -, n_j$. We will also denote the mean of the population receiving treatment j by μ_j and the variance of the population receiving treatment j by σ_j^2. Thus, we may say that the x_{ij} are independently and normally distributed with mean μ_j and variance σ^2.

Now consider the identity $x_{ij} = \mu + (\mu_j - \mu) + (x_{ij} - \mu_j)$. If we let $\beta_j = \mu_j - \mu$ and

$$\mu = \sum_{j=1}^{r} n_j \mu_j / N,$$

where

$$N = \sum_{j=1}^{r} n_j ,$$

then

$$\sum_{j=1}^{r} n_j B_j = \sum_{j=1}^{r} n_j (\mu_j - \mu) = \sum_{j=1}^{r} n_j \mu_j$$

$$- \sum_{j=1}^{r} n_j \mu = N \mu - N \mu = 0 .$$

If we also define $e_{ij} = x_{ij} - \mu_j$, then e_{ij} has mean 0 since the mean of x_{ij} is μ_j. Since the e_{ij} and x_{ij} differ by a constant they both have the same variance σ^2. Thus the assumptions 1, 2, and 3 given above may be written

$$x_{ij} = \mu + B_j + e_{ij}, \; i = 1, 2, ---, n_j, \; j = 1, 2, ---, 6.$$

e_{ij} are independently $N(0, \sigma^2)$

$$\sum_{j=1}^{r} n_j B_j = 0 \text{ (which reduces to } \sum_{j=1}^{r} B_j = 0 \text{ if all } n_j = n).$$

Now the null hypothesis and the alternate hypothesis may be written

H_0: $B_j = 0$, $j = 1, 2, ---, 6$

H_1: not all the B_j are zero.

Thus, each B_j is a measure of the deviation of the j^{th} population mean from the average of all six population means. If all 6 means are equal, then every B_j is zero.

The Research Design

The research design for the present study was based on The Solomon Four-Group Design. The comparison of achievement between two different review methods for a pre-calculus undergraduate mathematics course was the basic rationale for conducting this research.

The research was conducted on three sections of Mathematics 109, Algebra and Trigonometry, at Wisconsin State University at LaCrosse, Wisconsin during the fall semester of the 1965-1966 academic year. There were five sections of Mathematics 109 during the fall semester. Of the three sections on which the research was conducted, two were taught by the writer and the other by another professor in the department.

Each of these three sections was divided into two groups; an experimental group and a control group. This division was accomplished by the use of a table of random numbers. The three experimental groups had eleven review sessions during the semester. During these review sessions, the experimental group watched and listened to a tape recorder synchronized with a slide projector. The three control groups also had eleven review sessions during the semester. These review sessions were conducted by the writer using the same techniques that were normally used to teach the writer's two sections during the semester. The same material was covered by both types of review sessions.

In the experimental review sessions, the students were seated in an auditorium and listened to a tape recorder. A slide projector was synchronized with the tape recorder. At different times, a slide would be projected on a screen at the front of the auditorium. The material on the slide was considered by the writer to be very basic to the course and consisted of definitions, theorems, proofs of selected theorems, and examples of certain basic concepts presented in the course. The material reviewed by the tape recorder was the basic material presented during the regular class periods. Each of these eleven meetings was of about one hour in length. These meetings were conducted by an audio-visual technician with little formal training in mathematics.

In the control review sessions, the students were seated in a large classroom. The writer conducted these review sessions using the usual lecture techniques with formal and informal class discussion. During these sessions, the writer reviewed the same material that was reviewed during the experimental review sessions. These sessions were conducted at the same time, but in a different building, as the experimental review sessions.

At the beginning of the semester, the writer's large class and the other professor's class were given a pretest. The test which was given consisted of two of the Cooperative Mathematics Tests of the Educational Testing Service of Princeton, New Jersey. The tests which were given were the Algebra III test, Form A, and the Trigonometry test, Form A. Each of these tests was forty minutes in length. In order to determine the effect of this pretesting upon the posttest results, the control group and the experimental group in the writer's smaller class were each divided into two groups through the use of a table of random numbers. One of the experimental groups and one of the control groups in this smaller class were also given the pretest, while the other experimental group and the other control group in this smaller class were not given this pretest.

The Cooperative Mathematics Tests were prepared by the staff of the Educational Testing Service in cooperation with many well known mathematics teachers throughout the United States. This collective action produced pretests that were administered to a national sample of students in 1960. These pretests were then reviewed and intensively revised. The new pretests were again administered to a national sample. From these latter results it was determined that these revised pretests were valid measures of developed abilities and thus their content validity was acceptable. The writer and Professor X examined the Algebra III tests, Form A and Form B, and the Trigonometry tests, Form A and Form B, and compared their content with the material to be covered in Mathematics 109. From the examination and comparison, the writer and Professor X judged that the content of these tests was valid with respect to the course content and educational aims of Mathematics 109.

The internal consistency of the Cooperative Mathematics Tests was measured by the Educational Testing Service. These reliabilities were computed from random subsamples using the Kuder-Richardson Formula 20. The writer also computed reliabilities for each of the four tests mentioned above. These reliabilities were determined by computing a coefficient of correlation using the "odds-evens" method. In this method, the number of correct odd responses and the number of correct even responses on each test were correlated. These correlations were adjusted by the use of the Spearman-Brown prophecy formula. Both sets of reliabilities are given in TABLE II. From these two sets of reliabilities, the writer decided that these tests were internally consistent for the subject matter and the students tested.

In order to facilitate the discussion which follows, the following notation will be used. The control and experimental groups of the writer's large class will be denoted by C_2 and E_2, respectively. The control and experimental groups of the other professor's class will be denoted by C_1 and E_1, respectively. The group of students in the writer's smaller class who were in the experimental group and who were not given a pretest will be denoted by $E_{4,N}$. The group of students in the writer's smaller class who were in the experimental group and who were given a pretest will be denoted by $E_{4,P}$. The group of students in this class who were in the control group and who were not given a pretest will be denoted by $C_{4,N}$. The group of students in this class who were in the control group and who were given a pretest will be denoted by $C_{4,P}$.

TABLE II

COEFFICIENTS OF RELIABILITY FOR THE COOPERATIVE MATHEMATICS TESTS USED IN THE STUDY

Test	Reliability	
	Educational Testing Service*	Writer**
Algebra III, Form A	.84	.79
Algebra III, Form B	.80	.72
Trigonometry, Form A	.78	.77
Trigonometry, Form B	.80	.86

* Computed using the Kuder-Richardson Formula 20

** Computed using the "odds-evens" method, adjusted with the Spearman-Brown Prophecy Formula.

At the end of the semester, the students in these three classes were given a posttest. The test which was given was Form B of the Algebra III Test given in the pretest and Form B of the Trigonometry Test given in the pretest.

By considering the posttest scores for the students in the writer's small class, the main effects of experimentation, the main effect of pretesting, and the interaction of testing with experimentation was estimated. A simple 2 x 2 analysis of variance design, as given in TABLE III, using the posttest scores, was used for this estimation. The main effect of experimentation was estimated from the column means of TABLE III. The main effect of pretesting was estimated from the row means of TABLE III. The interaction of testing with experimentation was estimated from the cell means of TABLE III.

TABLE III

ANALYSIS OF VARIANCE FOR SELECTED MEANS

	Control	Experimental
Pretested	$C_{4,P}$	$E_{4,P}$
Unpretested	$C_{4,N}$	$E_{4,N}$

The review materials for both the control and experimental groups were developed by the writer. In order to estimate the applicability of this review technique for instructor's other than the writer, a 2 x 2 analysis of variance design, as given in TABLE IV, using the posttest scores, was used for this estimation. The main effects of experimentation were estimated from the column means of this table. The main effects of the particular instructor were estimated from the row means of this table. The interaction of the particular instructor and experimentation was estimated from the cell means of this table.

TABLE IV

ANALYSIS OF VARIANCE FOR MEANS

	Control	Experimental
Writer	C_2, $C_{4,N}$, $C_{4,P}$	E_2, $E_{4,N}$, $E_{4,P}$
Other Professor	C_1	E_1

This design was based upon The Solomon Four-Group Design. This design permitted estimation for the main effects of experimentation, for the main effects of pretesting, for the main effects of the instructor, for the main effects of the interaction of the instructor and experimentation, and the main effects of the interaction of pretesting and experimentation.

CHAPTER IV

ORGANIZATION AND ANALYSIS OF THE DATA

Introduction

The present study was concerned with the relationship between achievement in pre-calculus mathematics and two review techniques. In one of these review methods, the students, in a group, listened to a tape recorder that reviewed the material that had been presented during the particular unit being reviewed. The students also watched a screen on which slides were projected by a slide projector that was synchronized with the tape recorder. The students who reviewed using this method were called the experimental group. In the other review method, the students, also in a single group, listened to a review of the same material conducted by the writer. This presentation consisted of both lecture and formal and informal class discussion. These students, unlike those in the other group, were permitted to ask pertinent questions. This latter group was called the control group.

The basic design of the study followed The Solomon Four-Group Design as given by Campbell and Stanley.[1] The use of this design enabled the researcher to estimate more accurately for the main effect of pretesting and for the main effect of the researcher as the instructor.

---

[1]Donald T. Campbell and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," Handbook of Research on Teaching, ed. N. L. Gage, (Chicago, 1963), pp. 183-195.

These estimations and the comparison between achievement under the control method and under the experimental review method were generally accomplished through an examination of the pretest and posttest scores of the students in the three classes involved in this research. The tables and descriptive analysis of the data presented in this chapter indicate the significant findings concerning the achievement under the two different methods of review.

Statistical Treatment

The statistics employed in the analysis of the relationship between the variables in the present research were chi-square and Snedecor's F - ratio. In addition to these two statistics, the t - ratio and the statistic

$$\frac{2.3026}{C}\left[(N - r) \log s_p^2 - \sum_{j=1}^{r} (n_j - 1) \log s_j^2\right]$$

were also used.

Concerning the chi-square statistic, Van Dalen and Meyer state the following:

> The basic notion underlying the chi-square technique, stated in terms of the null hypothesis, is that the observed frequencies in a category are a chance departure from the hypothetical or expected frequencies for the category. These expected frequencies are derived from any definition one might want to give the null hypothesis - - -
>
> $$\chi^2 = \text{Sum of } \frac{(O - E)^2}{E}$$
>
> where O = observed frequency in the category
>
> E = expected frequency.[2]

[2]Debold B. Van Dalen and William J. Meyer, Understanding Educational Research, (New York, 1962), p. 330.

The F - ratio is defined as the ratio between two quotients. Each of the quotients is a chi-square value which is divided by its own number of degrees of freedom. Symbolically the F - ratio may be defined as

$$F = \frac{\chi_1^2 / df_1}{\chi_2^2 / df_2}$$

Lindquist notes that:

> It should be apparent from the definition of F that the ratio between the estimates $\sum (X - M)^2 (n - 1)$ of the population variance derived from two random samples drawn from the same normal population is distributed as F. Accordingly, given the variance estimates obtained from different populations, we may, on the assumption that the populations are normal, test the hypothesis that the populations have the same variance.[3]

The t - ratio is defined as the ratio between a randomly selected normal random variable expressed in units of the population standard deviation and the square root of a randomly selected chi-square divided by its degrees of freedom. If X is normally distributed for a population whose mean is μ and whose variance is σ^2, if $z = \frac{X - \mu}{\sigma}$, and if we select a z at random from this population and independently select a chi-square at random from the chi-square distribution for k degrees of freedom, then we may symbolically form a t - ratio as follows:

$$t = \frac{z}{\sqrt{\frac{\chi^2}{k}}}.$$

When we wish to test the hypothesis H_0 that two means are equal, the t - distribution may be used. Johnson and Jackson state that:

---

[3] E. F. Lindquist, "Design and Analysis of Experiments in Psychology and Education," (Cambridge, Mass., 1956), p. 40.

> For samples drawn from a normal population, therefore, we know the sampling distribution of t, and accordingly may use the table of t to test the hypotheses H_0 whenever it specifies, or we may assume that the sample we have actually observed has been drawn from a population normal, or reasonably normal, in form.[4]

The statistic

$$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r}(n_j - 1)\log s_j^2\right]$$

may be used to test the hypothesis that several variances are equal. This is frequently necessary in analysis of variance problems in which we might doubt that a number of population variances are equal. Guenther discusses this statistic in the following manner:

> Under the assumptions that (a) r random samples are drawn from r populations and (b) the r populations are normal, the statistic
>
> $$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r}(n_j - 1)\log s_j^2\right]$$
>
> is approximately distributed as chi-square with r - 1 degrees of freedom if H_0: $\sigma_1^2 = \sigma_2^2 = - - - = \sigma_r^2$ is true. Here $s_1^2, s_2^2, - - -, s_r^2$ are the r sample variances. The sample sizes are $n_1, n_2, - - -, n_r$ with
>
> $$\sum_{j=1}^{r} n_j = N.$$
>
> Also
>
> $$s_p^2 = \frac{\sum_{j=1}^{r}(n_j - 1)s_j^2}{N - r}$$
>
> and

[4]Palmer O. Johnson and Robert W. B. Jackson, Modern Statistical Methods: Descriptive and Inductive, (Chicago, 1959), p. 151.

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

The more the s_j^2's differ from one another, the larger this statistic becomes. If the s_j^2 are all nearly the same, then the statistic is small. Hence H_0 is rejected only for large values.[5]

These four statistics are the major ones used in the analysis of the data accumulated for this research study. In order for the reader to more fully understand the discussion presented on the following pages, the following definitions are given.

(a) A random variable is a variable quantity whose value is determined by the outcome of a random experiment.

(b) A population is a set or collection of observations.

(c) A sample is a subset or a part of a population.

(d) A parameter is a quantity that could be computed from a population if the entire population were available. The mean μ and the variance σ^2 are parameters.

(e) A statistic is a quantity computed from a sample. The sample mean $\bar{X}$ and the sample variance s^2 are statistics.

(f) A hypothesis is an assumption about the form of a population or its parameters.

(g) A null hypothesis is a hypothesis of no differences between the form of a population or its parameters.

(h) A test is a rule or procedure used for deciding whether to accept or reject the hypothesis.

(i) The critical region is the set of outcomes for the experiment which leads to the rejection of the hypothesis.

(j) A Type I error is committed when a true hypothesis is rejected.

(k) A Type II error is committed when a false hypothesis is accepted.

---

[5]William C. Guenther, Analysis of Variance, (Englewood Cliffs, 1964), pp. 20-21.

(l) The level of significance is the probability of committing a Type I error and will be denoted by the Greek letter α.

(m) The power of the test is the probability of rejecting the hypothesis.

(n) A random sample is a sample chosen from a finite population in such a way that every sample of the same size has an equal chance of being selected.

(o) An unbiased estimate of a parameter is a statistic whose average value is equal to the parameter. The sample mean $\bar{X}$ and the sample variance

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{X})^2 / (n - 1)$$

are unbiased estimates of the population mean μ and the population variance σ^2, respectively.

It was the objective of this research to determine if, at the .05 level of confidence, the observed frequency of the variables considered was a chance departure from the expected frequency for the given category.

In the discussion which follows, these assumptions were explicitly made:

(a) the number of random samples were drawn from the same number of populations;

(b) these populations were normal; and

(c) each of the populations had the same variance.

The latter assumption is made only when testing the hypothesis of equal means.

As stated previously, most of the students enrolled in Mathematics 109 were freshmen. Furthermore, most of them were first-semester freshmen. Practically all of these students pre-registered for the fall semester at four pre-registration days during the preceding July. When registering, these students merely indicated the course which they wished to take during the fall semester. They were given no opportunity to select instructors or sections. The section assignments were made by members of the Registrar's staff during the interval between the students registering and the first of September. There was no known pattern to this assignment. Therefore, it seemed reasonable to assume that the placement of the students in the different sections of Mathematics 109 was done on a random basis. Thus, it appeared reasonable to make the assumption that the three sections involved in this research were, in fact, normal populations with respect to the mathematical preparation and ability of students in attendance at Wisconsin State University, La Crosse.

Since the students in each of these sections was assigned to an experimental or control group through the use of a table of random numbers, it seems logical, from the foregoing discussion, to assume that C_1, E_1, C_2, E_2, $C_{4,P}$, $C_{4,N}$, $E_{4,P}$, and $E_{4,N}$ were each normal populations. This assumption will be used throughout the ensuing discussion. The pretest was given, not to check on the normality of these populations, but to furnish further information through the use of gain scores.

Analysis of Pretest Scores

With the three assumptions as stated previously, the null hypothesis that the means of the algebra pretest scores of C_1, E_1, C_2, E_2, $C_{4,P}$, and $E_{4,P}$ were all equal was true at the five per cent level of significance. That is H_0: ${}_a\overline{\overline{X}}_{C_1} = {}_a\overline{\overline{X}}_{E_1} = {}_a\overline{\overline{X}}_{C_2} = {}_a\overline{\overline{X}}_{E_2} = {}_a\overline{\overline{X}}_{C_{4,P}} = {}_a\overline{\overline{X}}_{E_{4,P}}$ was accepted for $\alpha = .05$.

TABLE V

ANALYSIS OF VARIANCE FOR THE MEANS OF THE ALGEBRA PRETEST SCORES

Source of Variation	SS*	d.f.**	MS ***	F****
Among Groups	172.6088	5	34.5218	1.32'
Within Groups	2,642.2697	101	26.1611	
Total	2,814.8785	106		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

In this notation, ${}_a\bar{X}_{C_1}$, denotes the mean for the algebra pretest scores of the control group for section one. Similar interpretations are to be given to the other notations. The computational results are summarized in TABLE V. Thus, if one makes the assumption that the samples were drawn randomly from normal populations with equal variances, the hypothesis that the means of the algebra pretest scores were the same was accepted at the five per cent level of significance. From TABLE V it is seen that $F_{5, 101} = 1.32$ for this study. In $F_{5, 101}$, the five represents the number of treatments less one. The one hundred one represents the total number of all the observations for all of the treatments less the number of treatments. Since 1.32 is less than $2.37 = F_{95; 5, 60}$, the acceptance of the hypothesis of equal means for the algebra pretest scores was permissible at the five per cent level of significance. This use of $F_{.95;\ 5,\ 60}$ rather than $F_{.95;\ 5,\ 101}$, which is not commonly found in tables of the F-distribution, is not uncommon. Lindquist points out that:

> ...the common procedure in practice is to use the F for the nearest combination of smaller degrees of freedom that can be found in the table.[6]

This procedure will be followed throughout the remainder of this chapter.

In accepting H_0 in the previous paragraph, the assumption was made that the variances for the algebra pretest scores of the different groups were equal. Using Bartlett's test we may test the hypothesis H_0 that all of the variances were equal against the hypothesis H_1 that at least two variances were different. In this test, the statistic

$$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r}(n_j - 1)\log s_j^2\right]$$

was used. This statistic is approximately distributed as chi-square with five degrees of freedom if H_0 is true. For this hypothesis, $N = 107$ and $r = 6$. The basic computations of this statistic for this hypothesis is given as follows:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r}\frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$C = \frac{151,477}{147,056} = 1.03006$$

$$s_p^2 = \sum_{j=1}^{r}(n_j - 1)\, s_j^2 \,/\, (N - r)$$

---

[6]E. F. Lindquist, "Design and Analysis of Experiments in Psychology and Education," (Cambridge, Mass., 1956), p. 39.

$$s_p^2 = \frac{2642.27051}{101} = 26.16109$$

$$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r}(n_j - 1)\log s_j^2\right]$$

$$= 2.2353(143.18366 - 140.26372)$$

$$= 2.2353(2.91994)$$

$$= 6.53$$

Since 6.53 is less than 11.07 = $\chi^2_{.95;\,5}$, the hypothesis of equal variances was accepted. This calculation reinforces the assumption that the variances of the various groups were equal.

TABLE VI

ANALYSIS OF VARIANCE FOR THE MEANS OF THE TRIGONOMETRY PRETEST SCORES

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	115.2699	5	23.0540	1.01'
Within Groups	2,316.8703	101	22.9393	
Total	2,432.1402	106		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

Under the three previous assumptions, the null hypothesis

$$H_0: {}_t\bar{X}_{C_1} = {}_t\bar{X}_{E_1} = {}_t\bar{X}_{C_2} = {}_t\bar{X}_{E_2} = {}_t\bar{X}_{C_{4,P}} = {}_t\bar{X}_{E_{4,P}}$$

was tested. Here again, the notation ${}_t\bar{X}_{C_1}$ is used to denote the mean for the trigonometry pretest scores of the control group for section one. Similar interpretations are also to be given to the other notation. The computations for the acceptance or rejection of this hypothesis is given in TABLE VI. The hypothesis of equal means would be rejected if $F_{5,\ 101}$ is less than $F_{.95;\ 5,\ 101}$. Since $F_{5,\ 101} = 1.01$ is less than $2.37 = F_{.95;\ 5,\ 60}$, the hypothesis of equal means was accepted at the five per cent level of significance.

TABLE VII

HYPOTHESES CONCERNING MEANS TESTED
USING PRETEST SCORES

Hypothesis	F - ratio
Equal means of the algebra pretest scores for all groups	1.32'
Equal means of the trigonometry pretest scores for all groups	1.01'

'Not significant at the 5% level

Again, the acceptance of this hypothesis of equal means for the trigonometry pretest scores was based on the assumption that the variances of the different groups were equal. Through the use of Bartlett's test, the hypothesis H_0 that all of the variances were equal was tested. For this hypothesis also, $N = 107$ and $r = 6$. The computations for the statistic used in this test were as follows:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N-r}\right]$$

$$= \frac{151,477}{147,056} = 1.03006$$

$$s_p^2 = \frac{\sum_{j=1}^{r} (n_j - 1)\, s_j^2}{N-r}$$

$$= \frac{2316.87033}{101} = 22.93931$$

$$\frac{2.3026}{C}\left[(N-r)\log s_p^2 - \sum_{j=1}^{r} (n_j - 1)\log s_j^2\right]$$

$$= 2.2353(137.41858 - 135.49325)$$

$$= 2.2353(1.92533)$$

$$= 4.30$$

Since 4.30 is less than 11.07 = $\chi^2_{.95}$, the hypothesis of equal variances was accepted. The assumption that the variances of the trigonometry pretest scores of these groups were equal was reinforced through these calculations.

TABLE VIII

HYPOTHESES CONCERNING VARIANCES TESTED USING PRETEST SCORES

Hypothesis	χ^2 -ratio
Equal variances of the algebra pretest scores for all groups	6.53'
Equal variances of the trigonometry pretest scores for all groups	4.30'

'Not significant at the 5% level

The results of these analyses are summarized in TABLE VII and TABLE VIII. These results indicate that these research groups had means and variances that were statistically equal. Since it had been assumed that these groups were a random sample from a normal population of students at Wisconsin State University, it was deemed advisable to carry out further analysis of the data and test other hypotheses.

Analysis of Posttest Scores

After a semester of experimentation on groups whose means and variances on the pretests were statistically equivalent at the five per cent level of significance, were the means and variances of the groups still statistically equivalent? That is, should the hypothesis, $H_0: {}_A\bar{X}_{C_1} = {}_A\bar{X}_{E_1} = {}_A\bar{X}_{C_2} = {}_A\bar{X}_{E_2} = {}_A\bar{X}_{C_{4,P}} = {}_A\bar{X}_{E_{4,P}} = {}_A\bar{X}_{C_{4,N}} = {}_A\bar{X}_{E_{4,N}}$, be accepted for $\alpha = .05$? The notation ${}_A\bar{X}_{C_1}$ denotes the mean for the algebra posttest scores of the control group for section one. The other notations are to be similarly interpreted. TABLE IX gives the summary of computational results. This table shows that $F_{7,\ 117} = 1.69$. The statistical acceptance of the hypothesis of equal means of the algebra posttest scores for the eight groups was possible at the five per cent level of significance since $F_{7,\ 117} = 1.69$ is less than $2.17 = F_{.95;\ 7,\ 60}$.

The acceptance of the hypothesis of equal means for the algebra posttest scores was based upon the assumption that the variances of the eight groups were equal. Again, the use of Bartlett's test permits a statistical evaluation of the hypothesis of equal variances. The basic computations for testing this hypothesis through the use of

Bartlett's test were as follows:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$= \frac{12526379}{9895088} = 1.26592$$

$$s_p^2 = \frac{\sum_{j=1}^{r} (n_j - 1)\, s_j^2}{N - r}$$

$$= \frac{2949.15796}{117} = 25.20648$$

$$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r}(n_j - 1)\log s_j^2\right]$$

$$= 1.81892(163.97667 - 158.94918)$$

$$= 1.81892(5.02749)$$

$$= 9.14$$

In these computations N = 125 and r = 8. Since 9.14 is less than 14.07 = $\chi^2_{.95;\ 7}$, the hypothesis of equal variances of the algebra posttest scores for these eight groups was accepted at the five per cent level of significance.

The hypothesis, H_0: ${}_T\bar{X}_{C_1} = {}_T\bar{X}_{E_1} = {}_T\bar{X}_{C_2} = {}_T\bar{X}_{E_2} = {}_T\bar{X}_{C_{4,P}} = {}_T\bar{X}_{E_{4,P}} = {}_T\bar{X}_{C_{4,N}} = {}_T\bar{X}_{E_{4,N}}$, was also tested. Here ${}_TX_{C_1}$ denotes the mean for the trigonometry posttest scores of the control group for section one. Similar interpretations are to be given to the other notations. The summary of the computational results for the testing of this hypothesis is given in TABLE X. From these computations, $F_{7,\ 117} = 1.26$. At the five per cent level of significance the hypothesis of equal means of the trigonometry posttest scores was accepted since $F_{7, 117} = 1.26$ is less than $2.17 = F_{.95; 7, 60}$.

TABLE IX

ANALYSIS OF VARIANCE FOR THE MEANS OF THE ALGEBRA POSTTEST SCORES

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	298.6690	7	42.6670	1.69'
Within Groups	2,948.1310	117	25.1977	
Total	3,246.8000	124		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

Bartlett's test was then used to determine statistically whether the assumption of equal variances of trigonometry posttest scores for the eight groups was warranted. In the following computations of the statistic used in Bartlett's test, N = 125 and r = 8.

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n-1} - \frac{1}{N-r}\right]$$

$$= \frac{12526379}{9895088} = 1.26592$$

$$s_p^2 = \frac{\sum_{j=1}^{r} (n_j - 1)\, s_j^2}{N - r}$$

$$s_p^2 = \frac{4185.29757}{117} = 35.77177$$

$$\frac{2.3026}{C}\left[(N - r) \log s_p^2 - \sum_{j=1}^{r} (n_j - 1) \log s_j^2 \right]$$

$$= 1.81892(181.76418 - 178.58471)$$

$$= 1.81892(3.17947)$$

$$= 5.78$$

Since 5.78 is less than $14.07 = \chi^2_{.95;\ 7}$, the hypothesis of equal variances of the trigonometry posttest scores for these eight groups was statistically accepted at the five per cent level of significance.

TABLE X

ANALYSIS OF VARIANCE FOR THE MEANS OF THE TRIGONOMETRY POSTTEST SCORES

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	355.0420	7	50.7203	1.26'
Within Groups	4,715.4060	117	40.3026	
Total	5,070.4480	124		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

The use of the experimental review technique gave results which did not differ statistically from the control review technique when this statistical comparison was made with respect to the means and the variances of the algebra and trigonometry posttest scores of the different groups. After the experimental treatment, there was no statistical difference at the five per cent level between the means and between the variances of the algebra and trigonometry posttest scores.

Next, a comparison of the posttest scores of $E_{4,N}$ and the posttest scores of $C_{4,N}$ was made. These scores would contain no interaction between pretesting and posttesting. If the results of the comparison verified the results already obtained, then this would lend some credence to the assumption that there was no interaction between pretesting and the experimentation. This would follow even though the number of students involved was small.

TABLE XI

ANALYSIS OF VARIANCE FOR THE MEANS OF THE ALGEBRA POSTTEST SCORES OF $E_{4,N}$ AND THE MEANS OF THE ALGEBRA POSTTEST SCORES OF $C_{4,N}$

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	38.6778	1	38.6778	1.37'
Within Groups	451.1000	16	28.1938	
Total	489.7778	17		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

The hypothesis, H_0: ${}_A\bar{X}_{E_{4,N}} = {}_A\bar{X}_{C_{4,N}}$, was then tested. The results are summarized in TABLE XI. Since $F_{1,\ 16} = 1.37$ is less than $4.49 = F_{.95;\ 1,\ 16}$, the hypothesis of equal means for the algebra posttest scores for the two groups $E_{4,N}$ and $C_{4,N}$ was accepted.

Bartlett's test was then used to test the hypothesis that the variance of the algebra posttest scores for the experimental group $E_{4,N}$ was equal to the variance of the algebra posttest scores for the control group $C_{4,N}$. The computation of the statistic used to test this hypothesis was as follows:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N-r}\right]$$

$$= \frac{3217}{3024} = 1.06382$$

$$s_p^2 = \sum_{j=1}^{r} (n_j - 1)\ s_j^2 \ /\!/ \ (N - r)$$

$$= 28.19375$$

$$\frac{2.3026}{C}\left[(N-r)\log s_p^2 - \sum_{j=1}^{r} (n_j - 1)\log s_j^2\right]$$

$$= 2.16446(23.20256 - 22.93672$$

$$= 2.16446(.26584)$$

$$= .58$$

Since .58 is less than $3.84 = \chi^2_{.95;\ 1}$, the hypothesis that these variances were equal was accepted at the five per cent level of significance. Thus, there was no statistical difference between the variances of the algebra posttest scores for the experimental and control groups which were not pretested.

The hypothesis that the mean of the trigonometry posttest scores of the experimental group $E_{4,N}$ was equal to the mean of the trigonometry posttest scores of the control group $C_{4,N}$ was then tested. TABLE XII summarizes the computational results. Since $F_{1, 16} = 1.05$ is less than $4.49 = F_{.95; 1, 16}$, this hypothesis was accepted for $\alpha = .05$. That is, there was no statistical difference between the means of the trigonometry posttest scores for the non-pretested experimental and control groups.

TABLE XII

ANALYSIS OF VARIANCE FOR THE MEANS OF THE TRIGONOMETRY POSTTEST SCORES OF $E_{4,N}$ AND THE MEANS OF THE TRIGONOMETRY POSTTEST SCORES OF $C_{4,N}$

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	55.2250	1	55.2250	1.05'
Within Groups	841.2750	16	52.5797	
Total	896.5000	17		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

Next, the hypothesis, $H_0: {}_T s^2_{E_{4,N}} = {}_T s^2_{C_{4,N}}$, was tested by using Bartlett's test. The results of the computation of the statistic used for this test are given on the following page:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N-r}\right]$$

$$= \frac{3217}{3024} = 1.06382$$

$$s_p^2 = \sum_{j=1}^{r} (n_j - 1)\, s_j^2 \,/\, (N - r)$$

$$= 52.57969$$

$$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r} (n_j - 1)\log s_j^2\right]$$

$$= 2.16446(27.53312 - 26.48424)$$

$$= 2.16446(1.04888)$$

$$= 2.27$$

Since 2.27 is less than $3.84 = \chi^2_{.95;\,1}$, the hypothesis that the variance of the trigonometry posttest scores for the experimental group $E_{4,N}$ was equal to the variance of the trigonometry posttest scores for the control group $C_{4,N}$ was accepted for $\alpha = .05$. Thus, there was also no statistical difference between the variances of the trigonometry posttest scores for the non-pretested experimental and control groups.

A comparison of the algebra posttest scores for the pretested experimental groups and the pretested control groups was then made. The hypothesis, H_0: ${}_A\bar{X}_{E_P} = {}_A\bar{X}_{C_P}$, was tested. Here, ${}_A\bar{X}_{E_P}$ denotes the mean of the algebra posttest scores for the pretested experimental group. The computations are summarized in TABLE XIII. Since $F_{1,\,105} = 0.00$ is less then $4.00 = F_{.95;\,1,\,60}$, the hypothesis that the mean of the algebra posttest scores of the pretested experimental group was equal to the mean of the algebra posttest scores of the pretested control group was accepted at the five per cent level of significance. Under the three basic assumptions, the means of the algebra pretest scores for these two groups were equal. The experimental treatment, then, did not result in any apparent statistical difference in the experimental group at the five per cent level of significance. That is, there was no significant difference in algebraic achievement between the two review methods.

TABLE XIII

ANALYSIS OF VARIANCE FOR THE MEANS OF THE ALGEBRA POSTTEST SCORES OF THE PRETESTED EXPERIMENTAL GROUP AND THE PRETESTED CONTROL GROUP

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	0.0014	1	0.0014	0.00'
Within Groups	2,748.1668	105	26.1730	
Total	2,748.1682	106		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

The test of the hypothesis, H_0: ${}_A s^2_{E_P} = {}_A s^2_{C_P}$, was then performed. The computations for the statistic used in Bartlett's test were as follows:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$C = \frac{291.927}{289.170} = 1.00953$$

$$s_p^2 = \sum_{j=1}^{r} (n_j - 1)\, s_j^2 / (N - r)$$

$$= 26.17301$$

$$\frac{2.3026}{C} \left[(N - r) \log s_p^2 - \sum_{j=1}^{r} (n_j - 1) \log s_j^2 \right]$$

$$= 2.28085(148.87425 - 146.01048)$$

$$= 2.28085(2.86377)$$

$$= 6.53$$

Since 6.53 is greater than $3.84 = \chi^2_{.95;\,1}$, the hypothesis that the variance of the algebra posttest scores for the three pretested experimental groups was equal to the variance of the algebra posttest scores for the three pretested control groups was rejected at the five per cent level of significance. This implies that the experimental treatment resulted in a change in the variability of the level of algebra achievement. An examination of TABLE XXIV reveals a greater variation in the control group. The calculations of the sample variances for these two groups helped to reinforce this conclusion. The sample variance for the three control groups C_1, C_2, and $C_{4,P}$, considered as a single sample, was calculated to be 34.85. In contrast, the sample variance of the sample formed by combining the three experimental groups E_1, E_2, and $E_{4,P}$ was found to be 16.98. Thus, there was a smaller variation in algebra achievement through the use of the experimental treatment.

TABLE XIV

ANALYSIS OF VARIANCE FOR THE MEANS OF THE TRIGONOMETRY POSTTEST SCORES OF THE PRETESTED EXPERIMENTAL GROUP AND THE PRETESTED CONTROL GROUP

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	2.2404	1	2.2404	0.06'
Within Groups	4,156.0587	105	39.5815	
Total	4,158.2991	106		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

Comparisons were also made of the trigonometry posttest scores for the pretested experimental groups and the pretested control groups. The hypothesis, $H_0: {}_T\bar{X}_{E_P} = {}_T\bar{X}_{C_P}$, was tested. The mean of the trigonometry posttest scores for the pretested experimental group was denoted by ${}_T\bar{X}_{E_P}$. TABLE XIV summarizes the results of the computations used in testing this hypothesis. Since $F_{1, 105} = 0.06$ is less than $4.00 = F_{.95; 1, 60}$, the hypothesis that the mean of the trigonometry posttest scores of the pretested experimental group was equal to the mean of the trigonometry posttest scores of the pretested control group was accepted at the five per cent level of significance. Since, under the original assumptions, the trigonometry pretest scores for these two groups were equal, the increase in trigonometry achievement was statistically the same for the two review methods.

The hypothesis, H_0: ${}_{T}s^2{}_{E_p} = {}_{T}s^2{}_{C_p}$, was then tested. The computations for the statistic used in Bartlett's test are summarized below:

$$C = 1 + \frac{1}{3(r - 1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$= \frac{291,927}{289,170} = 1.00953$$

$$s_p^2 = \sum_{j=1}^{r} (n_j - 1)\, s_j^2 \,/\, (N - r)$$

$$= 26.17301$$

$$\frac{2.3026}{C}\left[(N - r) \log s_p^2 - \sum_{j=1}^{r} (n_j - 1) \log s_j^2\right]$$

$$= 2.28085(167.73750 - 167.41959)$$

$$= 2.28085(.31791)$$

$$= .73$$

Since .73 is less than $3.84 = \chi^2_{.95;\,1}$, the hypothesis that the variance of the trigonometry posttest scores for the pretested experimental group was equal to the variance of the trigonometry posttest scores of the pretested control groups was accepted at the five per cent level of significance. Therefore, there was no statistically apparent change in the variability of trigonometry achievement using either the mechanical review device or the non-methanical review approach.

This research design was constructed to enable generalizations to be made concerning above normal-sized classes and different professors at Wisconsin State University, La Crosse. The next two comparisons were made in an attempt to secure more complete evidence which might help to make these generalizations. The first of these comparisons was concerned with the experimental and control groups of the above normal-sized class.

The hypothesis, H_0: ${}_A\bar{X}_{C_2} = {}_A\bar{X}_{E_2}$, was tested to aid in this generalization. The computations for the statistic used in testing this hypothesis are summarized in TABLE XV. Since $F_{1, 56} = 0.00$ is less than $4.08 = F_{.95; 1, 40}$, the hypothesis that the mean of the algebra posttest scores for the control group C_2 was equal to the mean of the algebra posttest scores for the experimental group E_2 was accepted at the five per cent level of significance.

TABLE XV

ANALYSIS OF VARIANCE FOR THE MEANS OF THE ALGEBRA POSTTEST SCORES FOR THE CONTROL GROUP C_2 AND THE ALGEBRA POSTTEST SCORES FOR THE EXPERIMENTAL GROUP E_2

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	0.0690	1	0.0690	0.00'
Within Groups	1,056.2069	56	18.8608	
Total	1,056.2759	57		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

The variances of these two groups were then compared. The hypothesis, H_0: ${}_As^2_{C_2} = {}_As^2_{E_2}$, was tested to aid in this comparison. The basic computations of the statistic used in testing this hypothesis are given below:

$$C = 1 + \frac{1}{3(r - 1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$= \frac{57}{56} = 1.01786$$

$$s_p^2 = \sum_{j=1}^{r} (n_j - 1)\, s_j^2 \,/\, (N - r)$$

$$= 18.86084$$

$$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r} (n_j - 1)\log s_j^2\right]$$

$$= 2.26220(71.43136 - 70.18760)$$

$$= 2.26220(1.24376)$$

$$= 2.81$$

Since 2.81 is less than $3.84 = \chi^2_{.95;\,1}$, the hypothesis that the variance of the algebra posttest scores for the control group C_2 was equal to the variance of the algebra posttest scores for the experimental group E_2 was accepted at the five per cent level of significance.

The next hypothesis to be tested was H_0: ${}_T\bar{X}_{C_2} = {}_T\bar{X}_{E_2}$. TABLE XVI summarizes the results of the computations for the statistic used in testing this hypothesis. Since $F_{1,\,56} = 0.06$ is less than $4.08 = F_{.95;\,1,\,40}$, the hypothesis that the mean of the trigonometry posttest scores for the control group C_2 was equal to the mean of the trigonometry posttest scores for the experimental group E_2 was accepted at the five per cent level of significance.

TABLE XVI

ANALYSIS OF VARIANCE FOR THE MEANS OF THE TRIGONOMETRY POSTTEST SCORES FOR THE CONTROL GROUP C_2 AND THE TRIGONOMETRY POSTTEST SCORES FOR THE EXPERIMENTAL GROUP E_2

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	2.0861	1	2.0861	0.06'
Within Groups	1,895.9311	56	33.8559	
Total	1,898.0172	57		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

The hypothesis, H_0: ${}_Ts^2_{C_2} = {}_Ts^2_{E_2}$, was then tested. Bartlett's test was again used to test this hypothesis. The summary of the computations for the statistic used in this test are given below:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N-r}\right]$$

$$= \frac{57}{56} = 1.01786$$

$$s_p^2 = \sum_{j=1}^{r} (n_j - 1)\, s_j^2 \,/\, (N - r)$$

$$= 33.85591$$

$$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r}(n_j - 1)\log s_j^2\right]$$

$$= 2.26220(85.66264 - 85.63632)$$

$$= 2.26220(.02632)$$

$$= .06$$

Since .06 is less than $3.84 = \chi^2_{.95;\,1}$, the hypothesis that the variance of the trigonometry posttest scores for the control group C_2 was equal to the variance of the trigonometry posttest scores for the experimental group E_2 was accepted at the five per cent level of significance.

These comparisons of the posttest scores of the control group C_2 and the experimental group E_2 reveal no statistical difference between the two groups. This evidence seems to indicate no appreciable difference between the experimental and the control group.

This next comparison will be an attempt to secure more evidence to permit a generalization concerning a different professor and the experimental treatment. The hypothesis, $H_0: {}_A\bar{\bar{X}}_{C_1} = {}_A\bar{\bar{X}}_{E_1}$, was tested to determine the effects, if any, of the other professor and the experimental treatment. The computations for the statistic used in testing this hypothesis are summarized in TABLE XVII. Since $F_{1,\,29} = 0.56$ is less than $4.18 = F_{.95;\,1,\,29}$, the hypothesis that the mean of the algebra posttest scores for the control group C_1 was equal to the mean of the algebra posttest scores for the experimental group E_1 was accepted at the five per cent level of significance.

TABLE XVII

ANALYSIS OF VARIANCE FOR THE MEANS OF THE ALGEBRA POSTTEST SCORES FOR THE CONTROL GROUP C_1 AND THE ALGEBRA POSTTEST SCORES FOR THE EXPERIMENTAL GROUP E_1

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	21.1473	1	21.1473	.56'
Within Groups	1,097.0462	29	37.8292	
Total	1,118.1935	30		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

To compare the variances of these two groups, the hypothesis, $H_0: {}_Ts^2_{C_1} = {}_Ts^2_{E_1}$, was tested. The computations for the statistic used in the test of this hypothesis are summarized below:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$= \frac{6243}{6032} = 1.03498$$

$$s_p^2 = \sum_{j=1}^{r} (n_j - 1)\, s_j^2 \;/\; (N - r)$$

$$= 37.82918$$

$$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r} (n_j - 1)\log s_j^2\right]$$

$= 2.22478(45.75707 - 44.47646)$

$= 2.22478(1.28061)$

$= 2.85$

Since 2.85 is less than $3.84 = \chi^2_{.95;\,1}$, the hypothesis that the variance of the algebra posttest scores for the control group C_1 was equal to the algebra posttest scores for the experimental group E_1 was accepted at the five per cent level of significance.

TABLE XVIII

ANALYSIS OF VARIANCE FOR THE MEANS OF THE TRIGONOMETRY POSTTEST SCORES FOR THE CONTROL GROUP C_1 AND THE TRIGONOMETRY POSTTEST SCORES FOR THE EXPERIMENTAL GROUP E_1

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	4.6890	1	4.6890	1.00'
Within Groups	1,357.3110	29	46.79348	
Total	1,362.0000	30		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

' Not significant at the 5% level

The hypothesis, H_0: ${}_T\bar{\bar{X}}_{C_1} = {}_T\bar{\bar{X}}_{E_1}$, was then tested. The results of the computations for the statistic used in testing this hypothesis are summarized in TABLE XVIII. Since $F_{1,\,29} = 1.00$ is less than $4.18 = F_{.95;\,1,\,29}$, the hypothesis that the mean of the trigonometry posttest scores for the control group C_1 was equal to the mean of the trigonometry posttest scores for the experimental group E_1 was accepted at the five per cent level of significance.

Finally, the variances of these two groups were compared. The hypothesis, H_0: ${}_Ts^2{}_{C_1} = {}_Ts^2{}_{E_1}$, was tested. The summary of the computations for the statistic used in this test was as follows:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$= \frac{6243}{6032} = 1.03498$$

$$s_p^2 = \sum_{j=1}^{r} (n_j - 1)\, s_j^2 \,/\, (N - r)$$

$$= 46.80383$$

$$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r} (n_j - 1)\log s_j^2\right]$$

$$= 2.22478(48.43812 - 48.33551)$$

$$= 2.22478(.10261)$$

$$= 0.23$$

Since 0.23 is less than $3.84 = \chi^2_{.95;\,1}$, the hypothesis that the variance of the trigonometry posttest scores for the control group C_1 was equal to the variance of the trigonometry posttest scores for the experimental group E_1 was accepted at the five per cent level of significance.

While these comparisons revealed no statistical difference between these two groups, an examination of TABLE XVII shows greater variation among groups for the algebraic achievement. This conclusion was reinforced by an examination of TABLE XXIV. This examination reveals a much greater variance for the control group C_1 than for the experimental group E_1. The conclusion here, then, was that Professor X's students using the control review method developed greater variation than using the experimental review method. This conclusion applies only to algebraic achievement.

TABLE XIX

HYPOTHESES CONCERNING MEANS TESTED USING POSTTEST SCORES

Hypothesis	F - ratio
Equal means of the algebra posttest scores for all groups	1.69'
Equal means of the trigonometry posttest scores for all groups	1.26'
Equal means of the algebra posttest scores for the non-pretested groups	1.37'
Equal means of the trigonometry posttest scores for the non-pretested groups	1.05'
Equal means of the algebra posttest scores for the pretested experimental and pretested control groups	0.00'
Equal means of the trigonometry posttest scores for the pretested experimental and pretested control groups	0.06'
Equal means of the algebra posttest scores for the large-sized class's experimental and control groups	0.00'
Equal means of the trigonometry posttest scores for the large-sized class's experimental and control groups	0.06'
Equal means of the algebra posttest scores for Professor X class's experimental and control groups	0.56'
Equal means of the trigonometry posttest scores for Professor X class's experimental and control groups	1.00'

' Not significant at the 5% level

TABLE XX

HYPOTHESES CONCERNING VARIANCES TESTED USING POSTTEST SCORES

Hypothesis	χ^2 - ratio
Equal variances of the algebra posttest scores for all groups	9.14'
Equal variances of the trigonometry posttest scores for all groups	5.78'
Equal variances of the algebra posttest scores for the non-pretested groups	0.58'
Equal variances of the trigonometry posttest scores for the non-pretested groups	2.27'
Equal variances of the algebra posttest scores for the pretested experimental and pretested control groups	6.53''
Equal variances of the trigonometry posttest scores for the pretested experimental and pretested control groups	0.73'
Equal variances of the algebra posttest scores for the large sized class's experimental and control groups	2.81'
Equal variances of the trigonometry posttest scroes for the large-sized class's experimental and control groups	0.06'
Equal variances of the algebra posttest scores for Professor X class's experimental and control groups	2.85'
Equal variances of the trigonometry posttest scores for Professor X class's experimental and control groups	0.23'

' Not significant at the 5% level
'' Significant at the 5% level

The results of the analyses using posttest scores are summarized in TABLE XIX and TABLE XX. These results show there was no significant statistical difference between the means and the variances of the posttest scores for the different groups involved in the study. From these results and from the results concerning pretest scores reported previously, the conclusion was drawn that neither the experimental review method nor the control review method resulted in any apparent significant statistical difference in either the level or the variation of algebraic or trigonometric achievement.

Analysis of Comparison of Pretest and Posttest Scores

The previous comparisons were concerned only with pretest or posttest scores. None of the hypotheses were concerned with any combination of pretest and posttest scores. In order to examine any increase in algebraic or trigonometric achievement, hypotheses concerning both pretest and posttest scores were tested. Since raw scores earned on different forms of a test are not directly comparable, the raw scores were replaced by converted scores. These converted scores were determined by the Educational Testing Service by taking raw scores on alternate forms of the algebra and trigonometry tests, equating them statistically, and converting them to a common score scale so that scores on both forms of the same test are comparable. In this section, all of the results have been obtained by the use of converted scores.

A comparison of the means for the pretest and posttest converted scores of both the Algebra III Test and the Trigonometry Test is given in TABLE XXI. In TABLE XXII, mid-percentile ranks are used rather than score means. TABLE XXIII uses percentile bands instead of mid-percentile ranks or mean scores. These mid-percentile ranks and percentile bands are based upon nationwide college norms developed by the Educational Testing Service. TABLE XXIV gives a comparison of the variances and the change in the variances of the different groups. From an examination of TABLES XXI, XXII, AND XXIII, there also appears to be no outstandingly significant differences between the experimental method and the control method in regards to achievement or variation in achievement.

TABLE XXI

A COMPARISON OF MEANS* FOR THE PRETEST AND POSTTEST CONVERTED SCORES

Group	Trigonometry (A)	Trigonometry (B)	Algebra (A)	Algebra (B)	(Gain) Trigonometry	(Gain) Algebra
C_1	144.29	151.65	143.71	151.24	7.36	7.53
E_1	139.57	152.93	143.71	154.14	13.36	10.43
Average	142.16	152.23	143.71	152.55	10.07	8.84
C_2	142.48	155.31	146.59	157.69	12.83	11.10
E_2	142.31	155.93	146.79	157.52	13.62	10.73
Average	142.40	155.62	146.69	157.60	13.22	10.91
$C_{4,P}$	146.22	161.11	147.22	158.56	14.89	11.34
$E_{4,P}$	140.56	154.00	140.67	153.00	13.44	12.33
Average	143.39	157.56	143.94	155.78	14.17	11.84
$C_{4,N}$		155.80		159.20		
$E_{4,N}$		150.50		154.25		
Average		153.44		157.00		
Composite Average	142.50	154.74	145.36	156.00		

* Rounded off to the nearest hundredth

TABLE XXII

A COMPARISON OF MID-PERCENTILE RANKS FOR THE PRETEST AND POSTTEST CONVERTED SCORES (BASED ON MEAN SCORES*)

Group	Trigonometry (A)	Trigonometry (B)	Algebra (A)	Algebra (B)	(Gain) Trigonometry	(Gain) Algebra
C_1	32	61	32	54	32 to 61	32 to 54
E_1	9	67	32	67	9 to 67	32 to 67
Average	25	67	32	61	25 to 67	32 to 61
C_2	25	72	38	83	25 to 72	38 to 83
E_2	25	77	38	83	25 to 77	38 to 83
Average	25	77	38	83	25 to 77	38 to 83
$C_{4,P}$	39	88	38	83	39 to 88	38 to 83
$E_{4,P}$	18	72	18	61	18 to 72	18 to 61
Average	25	81	32	73	25 to 81	32 to 73
$C_{4,N}$		77		83		
$E_{4,N}$		61		67		
Average		67		73		
Composite Average		72		73		

*Rounded off to the nearest integer

A comparison of the means of the algebra pretest and posttest converted scores for the experimental groups is given in TABLE XXV. As expected, the hypothesis that the means of these two groups of converted scores were equal was rejected. In testing this hypothesis,

$F_{1, 102}$ was computed to be 52.25. Since this was considerably greater than $F_{.95; 1, 60} = 4.00$, the hypothesis was rejected. Certainly, growth in achievement during the semester under the experimental treatment would be desirable, and to be expected.

TABLE XXIII

A COMPARISON OF PERCENTILE BANDS FOR THE PRETEST AND POSTTEST CONVERTED SCORES(BASED ON MEAN SCORES*)

Group	Trigonometry (A)	Trigonometry (B)	Algebra (A)	Algebra (B)
C_1	13-54	48-81	14-50	32-73
E_1	5-39	48-81	14-50	50-83
Average	9-48	48-81	14-50	44-78
C_2	9-48	54-85	22-54	67-91
E_2	9-48	61-88	22-54	67-91
Average	9-48	61-88	22-54	67-91
$C_{4,P}$	18-61	77-95	22-54	67-91
$E_{4,P}$	5-39	54-85	7-32	44-78
Average	9-48	67-94	14-50	54-87
$C_{4,N}$		61-88		67-91
$E_{4,N}$		39-77		50-83
Average		48-81		54-87
Composite Average		54-85		54-87

*Rounded off to the nearest integer

TABLE XXIV

A COMPARISON OF VARIANCES* FOR THE PRETEST AND POSTTEST CONVERTED SCORES

Group	Trigonometry (A)	Trigonometry (B)	Algebra (A)	Algebra (B)	(Change) Trigonometry	(Change) Algebra
C_1	38	115	66	138	+77	+72
E_1	45	89	53	59	+44	+ 6
Average	41	103	60	102	+62	+42
C_2	69	71	85	67	+ 2	-18
E_2	75	77	55	36	+ 2	-19
Average	73	74	70	52	+ 2	-18
$C_{4,P}$	58	110	58	59	+52	+ 1
$E_{4,P}$	41	53	18	58	+12	+40
Average	50	82	38	59	+32	+21
$C_{4,N}$		71		58		
$E_{4,N}$		175		99		
Average		117		76		
Composite Average		88		69		

*Rounded off to the nearest integer.

The inequality of the algebra pretest and posttest mean converted scores for the experimental groups does not imply that the variance of these two groups of converted test scores were equal. The hypothesis that these variances were equal was then tested using Bartlett's test.

The computation of the statistic used in this test for these groups of converted scores was as follows:

$$C = 1 + \frac{1}{3(r - 1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$= \frac{103}{102} = 1.00980$$

$$s^2 = \frac{\sum_{j=1}^{r} (n_j - 1)\, s_j^2}{N - r}$$

$$= \frac{78735}{1276} = 61.70455$$

$$\frac{2.3026}{C}\left[(N - r)\log s_p^2 - \sum_{j=1}^{r} (n_j - 1)\log s_j^2\right]$$

$$= 2.28024(182.61264 - 182.27961)$$

$$= 2.28024(.33303)$$

$$= .76$$

Since .76 is less than $3.84 = \chi^2_{.95,\ 1}$, the hypothesis of equal variances was accepted. Thus, the experimental treatment did not result in a change in the variances of the algebra test scores. This suggests that the variability in algebra achievement was not statistically affected by the use of the experimental treatment. This implies that, if one of the fifty-four students in the experimental group that took both a pretest and a posttest was selected at random, then the probability would be ninety-five per cent that the students' posttest converted algebra score would be no farther from the mean of the posttest converted algebra scores than the students' pretest converted algebra score was from the mean of the pretest converted algebra scores.

That is, the growth in algebraic achievement was no greater for those who scored high on the pretest than for those who scored low on the pretest. From this, there is the implication that the experimental review method was as effective for low achievers as for high achievers. It would appear that further examination in this area would be desirable.

TABLE XXV

ANALYSIS OF VARIANCE FOR THE MEANS OF THE ALGEBRA PRETEST AND POSTTEST CONVERTED SCORES FOR THE GROUPS E_1, E_2, AND $E_{4,P}$

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	3,102.1538	1	3,102.1538	52.25''
Within Groups	6,055.9616	102	59.3722	
Total	9,158.1154	103		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

'' Significant at the 5% level

The means of the trigonometry pretest and posttest converted scores for the three experimental groups, considered as a single group, are compared in TABLE XXVI. Again, the rejection of the hypothesis that the means of these two groups of converted scores were equal was to be expected. For testing this hypothesis, $F_{1, 102}$ was computed. Since $F_{1, 102} = 69.55$ is greater than $4.00 = F_{1, 60}$, the hypothesis was rejected. Growth in trigonometry achievement during the semester would be desirable. The experimental procedure would be nearly useless if no achievement was noted under this treatment.

TABLE XXVI

ANALYSIS OF VARIANCE FOR THE MEANS OF THE TRIGONOMETRY PRETEST AND POSTTEST CONVERTED SCORES FOR THE GROUPS E_1, E_2, AND E_4, P

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	4,751.0096	1	4,751.0096	69.55''
Within Groups	6,967.9039	102	68.3128	
Total	11,718.9135	103		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

'' Significant at the 5% level

The means of the trigonometry pretest and posttest converted scores for the experimental groups were unequal. This inequality of means does not imply that the variances of these two groups of converted test scores were equal. Bartlett's test was used to test the hypothesis that the variances of the trigonometry pretest and posttest converted scores for the experimental group were equal. The computation of the statistic used in this test for these groups of scores was as follows:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$= \frac{103}{102} = 1.00980$$

$$s_p^2 = \frac{\sum_{j=1}^{r} (n_j - 1)\, s_j^2}{N - r}$$

$$= 70.97943$$

$$\frac{2.3026}{C}\left[(N - r) \log s_p^2 - \sum_{j=1}^{r} (n_j - 1) \log s_j^2\right]$$

$$= 2.28024(188.81628 - 188.53119)$$

$$= 2.28024(.28509)$$

$$= .66$$

Since .66 is less than $3.84 = \chi^2_{.95;\, 1}$, the hypothesis of equal variances was accepted. That is, the variance of the trigonometry pretest converted scores of the experimental groups E_1, E_2, and $E_{4,P}$ was not statistically different from the variance of the trigonometry posttest converted scores of the same experimental groups at the five per cent level of significance. Here again, the experimental treatment did not result in any greater statistical variability in trigonometry achievement. Here again, this implies that the experimental review method was as effective for those scoring low on the trigonometry pretest as for those scoring high on the trigonometry pretest. Further examination in this area would also appear to be desirable.

This research study was designed to allow for testing the effect of pretesting. A comparison of the posttest converted scores of the non-pretested experimental group with the pretest converted scores of the control group consisting of the groups C_1, C_2, and $C_{4,P}$ was then made. The hypothesis H_0: ${}_a\bar{X}_{C_P} = {}_A\bar{X}_{E_{4,N}}$, was first tested. Here ${}_a\bar{X}_{C_P}$ denotes the mean of the algebra pretest converted scores of the control groups C_1, C_2, and $C_{4,P}$. The computations used for testing this hypothesis are summarized in TABLE XXVII. Since $F_{1, 61} = 6.49$ is greater than $4.00 = F_{.95; 1, 60}$, the hypothesis that the mean of the algebra pretest converted acores of the control group consisting of the control groups C_1, C_2, $C_{4,P}$ was equal to the mean of the algebra posttest converted scores of the experimental group $E_{4,N}$ was rejected at the five per cent level of significance. This conclusion, along with an examination of TABLE XXI, implies that the experimental treatment did result in an increase in algebra achievement.

TABLE XXVII

ANALYSIS OF VARIANCE FOR THE MEANS OF THE ALGEBRA PRETEST CONVERTED SCORES FOR THE CONTROL GROUPS C_1, C_2, AND $C_{4,P}$ AND THE ALGEBRA POSTTEST CONVERTED SCORES FOR THE EXPERIMENTAL GROUP $E_{4,N}$

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	498.6841	1	498.6841	6.49''
Within Groups	4,684.3001	61	76.7918	
Total	5,182.9842	62		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

'' Significant at the 5% level

A comparison of the variability of these two groups was then made. The hypothesis, $H_0: {}_{a}s^2{}_{C_P} = {}_{A}s^2{}_{E_{4,N}}$, was tested. The computations for the statistic used in Bartlett's test were as follows:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$= \frac{72,517}{69,174} = 1.04833$$

$$s^2 = \sum_{j=1}^{r} (n_j - 1)\, s_j^2 \,/\, (N - r)$$

$$= 75.15246$$

$$\frac{2.3026}{C}\left[(N - r) \log s_p^2 - \sum_{j=1}^{r} (n_j - 1) \log s_j^2\right]$$

$$= 2.19645(115.00391 - 114.88240)$$

$$= 2.19645(.12151)$$

$$= .27$$

Since .27 is less than $3.84 = \chi^2_{.95;\ 1}$, the hypothesis that the variance of the algebra pretest converted scores of the group consisting of the control groups C_1, C_2, and $C_{4,P}$ was equal to the variance of the algebra posttest converted scores of the experimental group $E_{4,N}$ was accepted at the five per cent level of significance.

Similarly, the hypothesis, H_0: ${}_t\bar{X}_{C_P} = {}_T\bar{X}_{E_{4,N}}$, was also tested. The mean of the trigonometry pretest converted scores of the control groups C_1, C_2, and $C_{4,P}$ was denoted by ${}_t\bar{X}_{C_P}$. TABLE XXVIII summarizes the results of the computations used for testing this hypothesis. Since $F_{1,\ 61} = 4.60$ is greater than $4.00 = F_{.95;\ 1,\ 60}$, the hypothesis that the mean of the trigonometry pretest converted scores for the control group consisting of C_1, C_2, and $C_{4,P}$ and the mean of the trigonometry posttest converted scores for the experimental group $E_{4,N}$ was rejected at the five per cent level of significance.

TABLE XXVIII

ANALYSIS OF VARIANCE FOR THE MEANS OF THE TRIGONOMETRY PRETEST CONVERTED SCORES FOR THE CONTROL GROUPS C_1, C_2, AND $C_{4,P}$ AND THE TRIGONOMETRY POSTTEST CONVERTED SCORES FOR THE EXPERIMENTAL GROUP $E_{4,N}$

Source of Variation	SS*	d.f.**	MS***	F****
Among Groups	327.2779	1	327.2779	4.60''
Within Groups	4,339.4364	61	71.1383	
Total	4,666.7143	62		

* Sum of Squares

** Degrees of Freedom

*** Mean Square

**** F - ratio

'' Significant at the 5% level

The hypothesis, $H_0: {}_t s^2_{C_P} = {}_T s^2_{E_{4,N}}$, was tested next. The computations for the statistic used in Bartlett's test of this hypothesis are given below:

$$C = 1 + \frac{1}{3(r-1)}\left[\sum_{j=1}^{r} \frac{1}{n_j - 1} - \frac{1}{N - r}\right]$$

$$= \frac{72,517}{69,174} = 1.04833$$

$$s_p^2 = \sum_{j=1}^{r} (n_j - 1)\, s_j^2 \,/\, (N - r)$$

$$= 70.85961$$

$$\frac{2.3026}{C}\left[(N - r) \log s_p^2 - \sum_{j=1}^{r} (n_j - 1) \log s_j^2\right]$$

$$= 2.19645(112.87440 - 110.65096)$$

$$= 2.19645(2.22344)$$

$$= 4.88$$

Since 4.88 is greater than $3.84 = \chi^2_{.95;\,1}$, the hypothesis that the variance of the trigonometry pretest converted scores of the control groups was equal to the variance of the trigonometry posttest converted scores of the experimental group $E_{4,N}$ was rejected at the five per cent level of significance. An examination of TABLE XXIV reveals that the variance of the experimental group $E_{4,N}$ was much greater than the variance of the pretested control group. One possible conclusion might be that the experimental treatment combined with the lack of any pretest "bias" resulted in increased variability. There is little, if any, evidence to support this conclusion.

TABLE XXIX

HYPOTHESES CONCERNING MEANS TESTED USING
PRETEST AND POSTTEST CONVERTED SCORES

Hypothesis	F - ratio
Equal means of the algebra pretest and posttest converted scores for the experimental groups	52.25''
Equal means of the trigonometry pretest and posttest converted scores for the experimental groups	69.55''
Equal means of the algebra pretest converted scores of the control groups and the posttest converted scores for the non-pretested experimental group	6.49''
Equal means of the trigonometry pretest converted scores of the control groups and the posttest converted scores for the non-pretested experimental group	4.60''

'' Significant at the 5% level

TABLE XXX

HYPOTHESES CONCERNING VARIANCES TESTED USING PRETEST AND POSTTEST CONVERTED SCORES

Hypothesis	χ^2 - ratio
Equal variances of the algebra pretest and posttest converted scores for the experimental groups	0.76'
Equal variances of the trigonometry pretest and posttest converted scores for the experimental groups	0.66'
Equal variances of the algebra pretest converted scores of the control groups and the posttest converted scores for the non-pretested experimental group	0.27'
Equal variances of the trigonometry pretest converted scores of the control groups and the posttest converted scores for the non-pretested experimental group	4.88''

' Not significant at the 5% level

'' Significant at the 5% level

The results in this section were obtained from comparisons of the pretest and posttest converted scores. The results are summarized in TABLE XXIX and TABLE XXX. From these results one can conclude that the experimental treatment did result in an increase in the level of both algebra and trigonometry achievement. One can also conclude that the use of the pretest did not significantly affect the posttest scores statistically. The experimental treatment did not result in any change in the variability of either algebra or trigonometry achievement. That is, the grouping of the scores around the mean remained the same under the experimental treatment. The final comparison in this section suggests, however, that the use of the pretest might have influenced the variability of trigonometry achievement. This might indicate a desirability for further study in this area.

Analysis of Comparisons Concerning Class Size and Instructor

All of the previous comparisons were concerned only with the effect of the experimental treatment. The last two of these comparisons were an attempt to develop more evidence to permit some generalization concerning the use of the experimental review method and larger than normal-sized classes or an instructor other than the writer.

The next comparisons were made in an attempt to estimate the interaction of testing and experimentation, the interaction of large classes and experimentation, and the interaction of different instructors and experimentation. The effects of these interactions were estimated by the use of a simple 2 x 2 analysis of variance.

For these comparisons, the assumptions which were made were the same as those previously given. These were that the four cells represented four random samples drawn from four populations, that each of the four populations was normal, and that each of the four populations had the same variance. It has been shown that these assumptions were plausible.

The first of these comparisons explored the effect of testing, the effect of experimentation, and the effect of interaction between testing and experimentation on algebra achievement. The three hypotheses that were tested in this comparison were:

1. H_0' : There was no difference between the means of the algebra posttest scores for the control review method and for the experimental review method.

2. H_0'' : There was no difference between the means of the algebra posttest scores for the pretested group and for the non-pretested group.

3. H_0''' : There was no interaction between the effects of experimentation and pretesting on algebra achievement. That is, the effects on algebra achievement of experimentation and pretesting were additive.

The results of the computations used to test these three hypotheses are summarized in TABLE XXXI.

TABLE XXXI

ANALYSIS OF VARIANCE OF ALGEBRA POSTTEST SCORES WITH MEANS ADJUSTED FOR DISPROPORTIONALITY

Source of Variation	Degrees of Freedom	Sum of Squares Unadjusted	Sum of Squares Adjusted*	Mean Square	F - ratio
Treatment	1	5.75515	6.07442	6.07442	.23'
Testing	1	8.85400	9.17327	9.17327	.35'
Interaction	1	47.53323	47.21396	47.21396	1.79'
Within	121	3,184.65762			
Total	124	3,246.80000			

* Adjustment term = -.31927

' Not significant at the 5% level

In this comparison, the frequencies for the four different subgroups were disproportional. This disproportionality was corrected by the technique discussed by Wert, Neidt, and Ahmann.

> When correcting disproportionality in a double classification with two categories within each classification - - - a simple and time-saving formula is available. For the purpose of developing the formula, a, b, c, and d, are designated as the frequencies in the four cells as in Table 81.

TABLE 81. SYMBOLICAL DESIGNATION
OF CELL FREQUENCIES IN A
FOUR-CELL TABLE

Stub Items	Headings		Total
	a	b	k_1
	c	d	k_2
Total	k_3	k_4	N

The mean score of k_1 cases is represented by $\bar{X}_1$, the mean score of k_2 cases by $\bar{X}_2$, the mean score of k_3 cases by $\bar{X}_3$, and the mean score of k_4 cases by $\bar{X}_4$. Furthermore, the difference between $\bar{X}_1$ and $\bar{X}_2$ equals $D_{1,\,2}$, whereas the difference between $\bar{X}_3$ and $\bar{X}_4$ equals $D_{3,4}$. The adjustment term for disproportion is equal to

$$\frac{\dfrac{(ad - bc)^2}{k_1k_2k_3k_4}\left[(k_1)(k_2)(D_{1,\,2})^2 + (k_3)(k_4)(D_{3,\,4})^2\right] - 2(D_{1,\,2})(D_{3,\,4})(ad - bc)}{N\left[1 - \dfrac{(ad - bc)^2}{k_1k_2k_3k_4}\right]}$$

This adjustment term, if positive, is to be subtracted from the sum of squares for interaction and added separately to the sum of squares for each of the two main effects, these sums of squares having been computed in the conventional manner. If negative, the adjustment term is added to the sum of squares for interaction and subtracted separately from the sums of squares for each of the two main effects.[7]

The F - ratios for a particular source of variation were computed by dividing the mean square for the source by the mean square of the

[7]James E. Wert, Charles O. Neidt, and J. Stanley Ahmann, Statistical Methods in Educational and Psychological Research, (New York, 1954), pp. 212-213.

source by the mean square of the variation within the groups. For the hypothesis H_0', the appropriate F - ratio was computed to be .23. Since this was less than $3.92 = F_{.95; 1, 120}$, the hypothesis that the mean of the algebra posttest scores for the control group was equal to the mean of the algebra posttest scores for the experimental group was accepted at the five per cent level of significance. For the hypothesis H_0'', the F - ratio was computed to be .35. Since this was less than $3.92 = F_{.95; 1, 20}$, the hypothesis that the mean of the algebra posttest scores for the pretested group was equal to the mean of the algebra posttest scores for the non-pretested group was accepted at the five per cent level of significance. For the last of these three hypotheses, H_0''', the F - ratio was computed to be 1.79. This was larger than the F - ratios computed for testing the first two of these three hypotheses. However, this value was less than $3.92 = F_{.95; 1, 120}$. Therefore, the hypothesis that there was no interaction between the effects of experimentation and pretesting on algebra achievement was accepted at the five per cent level of significance. Thus, the conclusion was drawn that pretesting had no significant effect upon algebra achievement as evidenced by the scores on the algebra posttest.

Since it was concluded that pretesting had no significant effect upon the algebra posttest converted scores, an analysis of covariance was then performed with the pretest converted scores being the covariate. In performing the analysis of covariance the following assumptions were made:

1. A random sample of size 1 was drawn from each of 107 populations;

2. each of the 107 populations was normal;

3. each of the 107 populations had the same variance;

4. the population means within each group lay on a straight line; and

5. the slope of the line was the same for each group.

The previous discussion in this chapter has shown that these assumptions were also reasonable.

Analysis of covariance was used to permit correction for initial differences in algebra achievement. Since a pretest had been given and since it was determined that the pretest had no effect upon the posttest converted scores, the writer decided to use the pretest converted scores for the covariate rather than some other criteria which might not be as immediately accessible as the pretest converted scores. The hypothesis tested was H_0: the "corrected" control review method effect was the same as the "corrected" experimental review method effect for algebra achievement. The results of the computations used to test this hypothesis are summarized in TABLE XXXIII. For this hypothesis, $F_{1, 104} = .47$ which was less than $4.00 = F_{.95; 1, 60}$. Therefore, the hypothesis that the "corrected" control review method effect was the same as the "corrected" experimental review method effect for algebra achievement was accepted at the five per cent level of significance. Thus, the conclusion was again made that there was no difference between the two review methods for achievement in algebra.

In the next comparison, the effect of testing, the effect of experimentation, and the effect of interaction between testing and experimentation on trigonometry achievement was examined. For this comparison, the three hypotheses that were tested were:

1. H_0' : There was no difference between the means of the trigonometry posttest scores for the control review method and for the experimental review method.

2. H_0'' : There was no difference between the means of the trigonometry posttest scores for the pretested group and the non-pretested group.

3. H_0''' : There was no interaction between the effects of experimentation and pretesting on trigonometry achievement. That is, the effects on trigonometry achievement of experimentation and pretesting were additive.

The results of the computations used to test these three hypotheses are summarized in TABLE XXXII.

TABLE XXXII

ANALYSIS OF VARIANCE OF TRIGONOMETRY POSTTEST SCORES WITH MEANS ADJUSTED FOR DISPROPORTIONALITY

Source of Variation	Degrees of Freedom	Sum of Squares Unadjusted	Sum of Squares Adjusted*	Mean Square	F - ratio
Treatment	1	16.60313	17.77207	17.77207	.43'
Testing	1	15.64893	16.81787	16.81787	.41'
Interaction	1	73.11426	71.94532	71.94532	1.75'
Within	121	4,965.08168		41.03373	
Total	124	5,070.44800			

* Adjustment term = 1.16894

' Not significant at the 5% level

Since the frequencies for the four different subgroups in this comparison were disproportional, this disproportionality was corrected using the same technique which was discussed previously. The F - ratios used to test these hypotheses were computed in the same manner as those used to test the similar hypotheses for algebraic achievement.

TABLE XXXIII

ANALYSIS OF COVARIANCE FOR THE ALGEBRA CONVERTED TEST SCORES

Source	SS_x[1]	SP[2]	SS_y[3]	SS'_y[4]	d.f.[5]	MS'_y[6]	F[7]
Treatments	21.46600	.22613	.00300	33.40367	1	33.40367	.47'
Error	6,653.31900	2,723.33462	8,466.97200	7,322.25416	104	70.40629	
Total	6,674.78500	2,723.56075	8,466.972	7,355.65783			

1. Sum of Squares for Pretest Converted Scores
2. Sum of Products
3. Sum of Squares for Posttest Converted Scores
4. Sum of Squares for Residuals
5. Degrees of Freedom
6. Mean Square
7. F - ratio

' Not significant at the 5% level

For this hypothesis H_0', the appropriate F - ratio was computed to be .43. Since this was less than $3.92 = F_{.95;\ 1,\ 120}$, the hypothesis that the mean of the trigonometry posttest scores for the control group was equal to the mean of the trigonometry posttest scores for the experimental group was accepted at the five per cent level of significance. For this hypothesis H_0'', the F - ratio was computed to be .41. Since this was less than $3.92 = F_{.95;\ 1,\ 120}$, the hypothesis that the mean of the trigonometry posttest scores for the pretested group was equal to the mean of the trigonometry posttest scores for the non-pretested group was accepted at the five per cent level of significance. For the last of these three hypotheses, H_0''', the F - ratio was computed to be 1.75. This was larger than the F - ratios computed for testing the first two of these hypotheses. However, since this value was less than $3.92 = F_{.95;\ 1,\ 120}$, the hypothesis that there was no interaction between the effects of experimentation and pretesting on trigonometry achievement was accepted at the five per cent level of significance. Therefore, it was again concluded that pretesting had no significant effect upon trigonometry achievement as evidenced by the scores on the trigonometry posttest.

Since the conclusion was made that pretesting had no significant effect upon the trigonometry posttest converted scores, an analysis of covariance was performed with the pretest converted scores being the covariate. Again, the use of the pretest converted scores as the covariate was dictated by the accessibility of the pretest converted scores. In performing this analysis of covariance, the assumptions were the same as those used for the previous analysis of covariance.

TABLE XXXIV

ANALYSIS OF COVARIANCE FOR THE TRIGONOMETRY CONVERTED TEST SCORES

Source	SS_x[1]	SP[2]	SS_y[3]	SS'_y[4]	d.f.[5]	MS'_y[6]	F[7]
Treatments	152.08100	21.60159	3.06800	39.59454	1	39.59454	.64'
Error	6,184.66700	4,066.37972	9,108.78200	6,435.16287	104	61.87656	
Total	6,336.74800	4,087.98131	9,111.85000	6,474.75741			

1. Sum of Squares for Pretest Converted Scores
2. Sum of Products
3. Sum of Squares for Posttest Converted Scores
4. Sum of Squares for Residuals
5. Degrees of Freedom
6. Mean Square
7. F - ratio

' Not significant at the 5% level

This analysis of covariance was used to permit correction for initial differences in trigonometry achievement. The hypothesis tested was H_0: the "corrected" control review method effect was the same as the "corrected" experimental review method effect for trigonometry achievement. TABLE XXXIV summarizes the results of the computations used to test this hypothesis. Since, for this hypothesis, $F_{1, 104} = .64$, is less than $4.00 = F_{.95; 1, 60}$, the hypothesis that the "corrected" control review method effect was the same as the "corrected" experimental review method effect for trigonometry achievement was accepted at the five per cent level of significance. Therefore, it was again concluded that there was no difference in trigonometry achievement between the two review methods.

From the preceding analyses, the conclusion was drawn that there was no significant difference between the regular review method and the review method using the mechanical device for students at Wisconsin State University, La Crosse. The design of this research study also permitted a comparison between experimentation and different professors.

The next group of analyses was concerned with estimating the effect of experimentation, the effect of different professors, and the effect of interaction between experimentation and different professors. The same assumptions that were made in estimating the interaction of testing and experimentation were used for these analyses.

The first of these comparisons explored the effect of experimentation and different professors upon algebra achievement. The three hypotheses tested in this comparison were:

1. H_0' : There was no difference between the means of the algebra posttest scores for the control review method and for the experimental review method.

2. H_0'': There was no difference between the means of the algebra posttest scores for the group taught by the writer and for the group taught by Professor X.

3. H_0''': There was no interaction between the effects of experimentation and of different professors on algebra achievement. That is, these effects on algebra achievement were additive.

TABLE XXXV summarizes the results of the computations used to test these three hypotheses. The same techniques were used for testing these hypotheses that were used in testing the interaction between experimentation and pretesting.

TABLE XXXV

ANALYSIS OF VARIANCE OF ALGEBRA POSTTEST SCORES WITH MEANS ADJUSTED FOR DISPROPORTIONALITY

Source of Variation	Degrees of Freedom	Sum of Squares Unadjusted	Sum of Squares Adjusted*	Mean Square	F - ratio
Treatment	1	5.75515	8.04194	8.04194	.34'
Professor	1	178.12773	180.41452	180.41452	7.72''
Interaction	1	234.03548	231.74869	231.74869	9.91''
Within	121	2,828.88164		23.379.19	
Total	124	3,246.80000			

* Adjustment term is 2.28679

' Not significant at the 5% level

'' Significant at the 5% level

For the hypothesis H_0', the computed F - ratio was .34. Thus the redundant conclusion was made that the mean of the algebra posttest scores for the control review method was equal to the mean of the algebra posttest scores for the experimental review method. The testing, and acceptance, of H_0' developed no new evidence. However, in testing H_0'', an F - ratio of 7.72 was computed. Since $F_{1,\ 12} = 7.72$ is greater than $F_{.95;\ 1,\ 120} = 3.92$, the hypothesis H_0'' was rejected. Thus, it was concluded that there was a difference between the mean of the algebra posttest scores for the group taught by the writer and the mean of the algebra posttest scores for the group taught by Professor X.

New evidence was also discovered in testing the hypothesis H_0'''. For this hypothesis, an F - ratio of 9.91 was computed. Since $F_{1,\ 121} = 9.91$ is greater than $F_{.95;\ 1,\ 120} = 3.92$, this hypothesis H_0''' was also rejected. That is, the hypothesis that there was no interaction between the effects of experimentation and of different professors on algebraic achievement was rejected at the five per cent level of significance.

These latter two conclusions appeared to indicate a bias between the writer and both of the review methods. However, it was possible that there was no significant difference in <u>gain</u> in algebraic achievement between the two instructors. In order to determine the plausibility of this conclusion, an analysis of variance of algebra converted gain scores was developed.

Using these gain scores for algebraic achievement, the three hypotheses that were tested were:

1. H_0' : There was no difference between the means of the algebra converted gain scores for the control review method and for the experimental review method.
2. H_0'' : There was no difference between the means of the algebra converted gain scores for the group taught by the writer and for the group taught by Professor X.
3. H_0''' : There was no interaction between the effects of experimentation and different instructors on algebra achievement as evidenced by algebra converted gain scores.

The results of the computations used to test these three hypotheses are summarized in TABLE XXXVI.

TABLE XXXVI

ANALYSIS OF VARIANCE OF ALGEBRA GAIN SCORES WITH MEANS ADJUSTED FOR DISPROPORTIONALITY

Source of Variation	Degrees of Freedom	Sum of Squares Unadjusted	Sum of Squares Adjusted*	Mean Square	F - ratio
Treatment	1	21.01592	25.09209	25.09209	.52'
Professor	1	115.75775	119.83392	119.83392	2.50'
Interaction	1	180.34007	176.26390	176.26390	3.68'
Within	103	4,937.52177		47.93710	
Total	106	5,254.63551			

* Adjustment term is -4.07617

' Not significant at the 5% level

For the hypothesis H_0', the computed F - ratio was .52. Since $F_{1, 103} = .52$ is less than $F_{.95; 1, 60} = 4.00$, the hypothesis that the mean of the algebra converted gain scores for the pretested control group was equal to the mean of the algebra converted gain scores for the pretested experimental group was accepted at the five per cent level of significance. This result reinforced the previous conclusion that the two review methods were equally effective in producing growth in algebraic achievement.

For the hypothesis H_0'', the computed F - ratio was 2.50. Since $F_{1, 103} = 2.50$ is less than $F_{.95; 1, 60} = 4.00$, the hypothesis that the mean of the algebra converted gain scores for the pretested group taught by the writer was equal to the mean of the algebra converted gain scores for the pretested group taught by Professor X was accepted at the five per cent level of significance. From this result and an examination of TABLE XXI, the conclusion was drawn that there was actually no statistical difference between the two professors in producing growth in algebraic achievement in their students at Wisconsin State University, La Crosse.

For the hypothesis H_0''', the computed F - ratio was 3.68. Since $F_{1, 103} = 3.68$ is less than $F_{.95; 1, 60} = 4.00$, the hypothesis that there was no interaction between the effects of experimentation and different instructors on growth in algebra achievement as evidenced by algebra gain scores was accepted at the five per cent level of significance. From this result, it was concluded that there was no apparent statistical interaction between experimentation and different instructors for algebra achievement. This conclusion was drawn on the assumption that increase in achievement was more desirable than the attainment of a specific achievement score.

The next comparisons attempted to estimate the effect of experimentation and different professors upon trigonometry achievement. The three hypotheses tested were:

1. H_0' : There was no difference between the means of the trigonometry posttest scores for the control review method and for the experimental review method.

2. H_0'' : There was no difference between the means of the trigonometry posttest scores for the group taught by the writer and for the group taught by Professor X.

3. H_0''' : There was no interaction between the effects of experimentation and of different professors on trigonometry achievement.

TABLE XXXVII summarizes the results of the computations used to test these three hypotheses.

TABLE XXXVII

ANALYSIS OF VARIANCE OF TRIGONOMETRY POSTTEST SCORES WITH MEANS ADJUSTED FOR DISPROPORTIONALITY

Source of Variation	Degrees of Freedom	Sum of Squares Unadjusted	Sum of Squares Adjusted*	Mean Square	F - ratio
Treatment	1	16.60313	19.66286	19.66286	.50'
Professor	1	118.57566	121.63539	121.63539	3.08'
Interaction	1	163.50574	160.44601	160.44601	4.07''
Within	121	4,771.76347		39.43606	
Total	124	5,070.44800			

* Adjustment term is 3.05973

' Not significant at the 5% level

'' Significant at the 5% level

For this hypothesis H_0', the computed F - ratio was .50. Since $F_{1,\ 121} = .50$ is less than $F_{.95;\ 1,\ 120} = 4.00$, the redundant conclusion was made that the mean of the trigonometry posttest scores for the control review method was equal to the mean of the trigonometry posttest scores for the experimental review method. Here also, the testing and acceptance of H_0' developed no new evidence. This procedure merely confirmed a previous conclusion.

For the hypothesis H_0'', the computed F - ratio was 3.08. Since $F_{1,\ 121} = 3.08$ is less than $F_{.95;\ 1,\ 120} = 4.00$, the hypothesis that the mean of the trigonometry posttest scores for the group taught by the writer was equal to the mean of the trigonometry posttest scores for the group taught by Professor X was accepted at the five per cent level of significance. For the hypothesis H_0''', the computed F - ratio was found to be 4.07. Since $F_{1, 121} = 4.07$ is greater than $F_{.95; 1, 120} = 4.00$, the hypothesis that there was no interaction between the effects of experimentation and of different professors on trigonometry achievement was rejected at the five per cent level of significance. This conclusion and the relatively large value of the F - ratio computed for the hypothesis H_0'' prompted the decision to examine the trigonometry converted gain scores with another analysis of variance.

Using the converted gain scores for trigonometry achievement, the three hypotheses tested were:

1. H_0' : There was no difference between the means of the trigonometry converted gain scores for the control review method and for the experimental review method.
2. H_0'' : There was no difference between the means of the trigonometry converted gain scores for the group taught by the writer and the group taught by Professor X.
3. H_0''' : There was no interaction between the effects of experimentation and different instructors on trigonometry achievement as evidenced by trigonometry converted gain scores.

The results of the computations used to test these three hypotheses are summarized in TABLE XXXVIII.

For the hypothesis H_0', the computed F - ratio was 1.94. Thus $F_{1, 103} = 1.94$ is less than $F_{.95; 1, 60} = 4.00$ and the hypothesis that the means of the trigonometry converted gain scores for the control review method was equal to the mean of the trigonometry converted gain scores for the experimental review method was accepted at the five per cent level of significance. It was noted, however, that the value of this F - ratio was considerably larger than that of any of the corresponding F - ratios computed previously. It appeared then that there was a greater variation between the two review methods for trigonometry achievement than for algebra achievement.

TABLE XXXVIII

ANALYSIS OF VARIANCE OF TRIGONOMETRY CONVERTED GAIN SCORES WITH MEANS ADJUSTED FOR DISPROPORTIONALITY

Source of Variation	Degrees of Freedom	Sum of Squares Unadjusted	Sum of Squares Adjusted*	Mean Square	F - ratio
Treatment	1	107.77744	121.84220	121.84220	1.94'
Professor	1	261.67782	275.74258	275.74258	4.39''
Interaction	1	526.75410	512.68934	512.68934	8.17''
Within	103	6,464.25793		62.75979	
Total	106	7,360.46729			

* Adjustment term is -14.06476

' Not significant at the 5% level

'' Significant at the 5% level

For the hypothesis H_0'', the computed F - ratio was 4.39. Since $F_{1,\ 103} = 4.39$ is greater than $F_{.95;\ 1,\ 60} = 4.00$, the hypothesis that the mean of the trigonometry converted gain scores for the pretested group taught by the writer was equal to the mean of the trigonometry converted gain scores for the pretested group taught by Professor X was rejected at the five per cent level of significance. The rejection of this hypothesis appears to contradict the previously accepted hypothesis that there was no difference between the means of the trigonometry posttest scores for the group taught by the writer and by the group taught by Professor X. However, this hypothesis was based solely on results from pretested students, while the previous hypothesis was based on results from posttested students. Thus, it appeared that groups $C_{4,N}$ and $E_{4,N}$ had an influence upon the previous hypothesis. From these results, it appeared that further study in this area would be desirable.

For the hypothesis H_0''', the computed F - ratio was 8.17. Since $F_{1, 103} = 8.17$ is greater than $F_{.95; 1, 60} = 4.00$, the hypothesis that there was no interaction between the effects of experimentation and different instructors on trigonometry achievement was rejected at the five per cent level of significance. The rejection of this hypothesis and the rejection of the previous hypothesis concerning interaction between experimentation and instructor based upon a comparison of posttest scores resulted in the conclusion that instructor bias was introduced into the research study.

An examination of TABLE XXI relealed that the bias which was introduced into the study was in favor of the writer for trigonometry achievement. One possible explanation of this apparent bias was the fact that the writer in his teaching as well as in both review methods developed the trigonometric functions from a "circular function" approach. Professor X might have used a more traditional "right-triangle" development of the trigonometric functions. Nevertheless, it appears that further study in this area would be desirable.

The design of this research study also permitted a comparison between experimentation and class size. The final set of analyses was concerned with estimating the effect of experimentationl the effect of class size, and the effect of interaction between experimentation and class size. Again, the same assumptions that were made in estimating the interaction of experimentation and testing were used for these analyses.

TABLE XXXIX

ANALYSIS OF VARIANCE OF ALGEBRA POSTTEST SCORES WITH MEANS ADJUSTED FOR DISPROPORTIONALITY

Source of Variation	Degrees of Freedom	Sum of Squares Unadjusted	Sum of Squares Adjusted*	Mean Square	F - ratio
Treatment	1	5.75515	7.62982	7.62982	.30'
Class size	1	94.07638	95.95105	95.95105	3.82'
Interaction	1	106.62536	104.75069	104.75069	4.17''
Within	121	3,040.34311		25.12680	
Total	124	3,246.80000			

* Adjustment term is 1.87467

' Not significant at the 5% level

'' Significant at the 5% level

The first analysis explored the effect of experimentation and class size upon algebra achievement. The hypotheses tested in this analysis were:

1. H_0' : There was no difference between the means of the algebra posttest scores for the control review method and for the experimental review method.

2. H_0'' : There was no difference between the means of the algebra posttest scores for the above normal-sized class and for the normal-sized classes.

3. H_0''' : There was no interaction between the effects of experimentation and of class size on algebra achievement.

TABLE XXXIX summarizes the results of the computations used to test these three hypotheses. The same techniques were also used for testing these hypotheses that were used in testing the interaction between experimentation and pretesting.

For the hypothesis H_0', the computed F - ratio was .30. Since $F_{1,\ 121} = .30$ is less than $F_{.95;\ 1,\ 120} = 3.92$, the hypothesis that the mean of the algebra posttest scores for the control review group was equal to the mean of the algebra posttest scores for the experimental review group was accepted at the five per cent level of significance. The testing of this hypothesis H_0' developed no new evidence. However, the acceptance of this hypothesis again confirmed the conclusion that there was no difference in either review method with respect to achievement in algebra.

For the hypothesis H_0'', the computed F - ratio was 3.82. Since $F_{1,\ 121} = 3.82$ is less than $F_{.95;\ 1,\ 120} = 3.92$, the hypothesis H_0'' was statistically accepted at the five per cent level of significance. Thus, it was concluded that the mean of the algebra posttest scores for the above normal-sized classes was equal to the mean of the algebra posttest scores for the normal-sized classes. However, since 3.82 was relatively close to 4.00, further examination was deemed advisable.

For the hypothesis H_0''', the computed F - ratio was 4.17. Since $F_{1,\ 121} = 4.17$ is greater than $F_{.95;\ 1,\ 120} = 3.92$, this hypothesis was rejected. Thus, it was concluded that there was interaction between the effects of experimentation and the effects of class size.

This last conclusion appeared to indicate a non-linear functional relationship between review method and the class sizes used in this investigation. However, it was also possible that there was no significant difference in <u>gain</u> in achievement in algebra between the two different sized classes. In order to determine the statistical correctness of this conclusion, another analysis of variance of algebra converted gain scores was developed.

The three hypotheses tested using the algebra gain scores were:

1. H_0' : There was no difference between the means of the algebra converted gain scores for the control review method and for the experimental review method.

2. H_0'' : There was no difference between the means of the algebra converted gain scores for the above normal-sized class and the normal-sized classes.

3. H_0''' : There was no interaction between the effects of experimentation and class size on algebra achievement as evidenced by algebra converted gain scores.

The results of the computations used to test these three hypotheses are summarized in TABLE XL.

TABLE XL

ANALYSIS OF VARIANCE OF ALGEBRA CONVERTED GAIN SCORES
WITH MEANS ADJUSTED FOR DISPROPORTIONALITY

Source of Variation	Degrees of Freedom	Sum of Squares		Mean Square	F - ratio
		Unadjusted	Adjusted*		
Treatment	1	21.01592	22.38010	22.38010	.45'
Class size	1	25.25021	26.61439	26.61439	.54'
Interaction	1	93.46379	92.09961	92.09961	1.85'
Within	103	5,114.90559		49.65928	
Total	106	5,254.63551			

* Adjustment term is -1.36418

' Not significant at the 5% level

For the hypothesis H_0', the computed F - ratio was .38. Since $F_{1,\ 103} = .38$ is less than $F_{.95;\ 1,\ 60} = 4.00$, the hypothesis that the mean of the algebra converted gain scores for the control review group was equal to the mean of the algebra converted gain scores for the experimental review group was accepted at the five per cent level of significance. This result confirmed the previous conclusion that the two review methods were equivalent in producing achievement in algebra.

For the hypothesis H_0'', the computed F - ratio was .54. Since $F_{1,\ 103} = .54$ is less than $F_{.95;\ 1,\ 60} = 4.00$, the hypothesis that the mean of the algebra converted gain scores for the above normal-sized class was equal to the mean of the algebra converted gain scores for the pretested students in the normal-sized classes was accepted at the five per cent level of significance. This result confirmed the acceptance of the previous hypothesis H_0'' which also concluded that there was no difference between algebra achievement and class size.

For the hypothesis H_0''', the computed F - ratio was 1.85. Since $F_{1,\ 103} = 1.85$ is less than $F_{.95;\ 1,\ 60} = 4.00$, the hypothesis that there was no interaction between the effects of experimentation and class size on algebra achievement as evidenced by algebra converted gain scores was accepted at the five per cent level of significance. This result contradicted the previous conclusion which was based upon posttest scores rather than converted gain scores. On the assumption that increase in each student's individual achievement was more desirable than the attainment of an arbitrary level of achievement for the entire class, the conclusion was drawn that there was no interaction between experimentation and class size for algebra achievement.

The final comparisons explored the effect of experimentation and class size upon trigonometry achievement. The hypotheses tested were:

1. H_0' : There was no difference between the means of the trigonometry posttest scores for the control review method and for the experimental review method.

2. H_0'' : There was no difference between the means of the trigonometry posttest scores for the above normal-sized class and for the normal-sized classes.

3. H_0''' : There was no interaction between the effects of experimentation and of class size on trigonometry achievement.

TABLE XLI summarizes the results of the computations used to test these hypotheses. The testing of these hypotheses used the same techniques that were previously used in testing the interaction between experimentation and pretesting.

TABLE XLI

ANALYSIS OF VARIANCE OF TRIGONOMETRY POSTTEST SCORES WITH MEANS ADJUSTED FOR DISPROPORTIONALITY

Source of Variation	Degrees of Freedom	Sum of Squares Unadjusted	Sum of Squares Adjusted*	Mean Square	F - ratio
Treatment	1	16.60313	18.56789	18.56789	.46'
Class size	1	43.05763	45.02239	45.02239	1.11'
Interaction	1	93.09761	91.13285	91.13285	2.24'
Within	121	4,917.68963		40.64206	
Total	124	5,070.44800			

* Adjustment term is 1.96476

' Not significant at the 5% level

For the hypothesis H_0', the computed F - ratio was .46. Since $F_{1,\ 121} = .46$ is less than $F_{.95;\ 1,\ 120} = 3.92$, the hypothesis that the mean of the trigonometry posttest scores for the control review group was equal to the mean of the trigonometry posttest scores for the experimental review group was accepted at the five per cent level of significance. From this, and previous conclusions, the decision was made that there was no statistical difference between the results obtained by either review method for both achievement in trigonometry and achievement in algebra.

For the hypothesis H_0'', the computed F - ratio was 1.11. Since $F_{1,\ 121} = 1.11$ is less than $F_{.95;\ 1,\ 120} = 3.92$, the hypothesis that the mean of the trigonometry posttest scores for the above normal-sized class was equal to the mean of the trigonometry posttest scores for the normal-sized class was accepted at the five per cent level of significance. That is, there appeared to be no statistical difference between the class size used and achievement in trigonometry.

For the hypothesis H_0''', the computed F - ratio was 2.24. Since $F_{1,\ 121} = 2.24$ is less than $F_{.95;\ 1,\ 120} = 3.92$, the hypothesis that there was no interaction between the effects of experimentation and of class size on trigonometry achievement was accepted at the five per cent level of significance. That is, the conclusion was drawn that the functional relationship between the variable denoted as experimentation and the variable denoted as class size was linear.

Even though the three preceding hypotheses were accepted, since a comparison of algebra converted gain scores was made estimating the effect of experimentation, the effect of class size, and the effect of interaction between experimentation and class size, a similar comparison was also made using trigonometry converted gain scores.

Using the converted gain scores for trigonometry achievement, the three hypotheses tested were:

1. H_0' : There was no difference between the means of the trigonometry converted gain scores for the control review method and for the experimental review method.

2. H_0'' : There was no difference between the means of the trigonometry converted gain scores of the above normal-sized class and for the normal-sized classes.

3. H_0''' : There was no interaction between the effects of experimentation and of class size on trigonometry achievement as evidenced by trigonometry converted gain scores.

The results of the computations used to test these three hypotheses are summarized in TABLE XLII.

TABLE XLII

ANALYSIS OF VARIANCE OF TRIGONOMETRY CONVERTED GAIN SCORES WITH MEANS ADJUSTED FOR DISPROPORTIONALITY

Source of Variation	Degrees of Freedom	Sum of Squares Unadjusted	Sum of Squares Adjusted*	Mean Square	F - ratio
Treatment	1	107.77744	113.14143	113.14143	1.68'
Class size	1	76.17699	81.54098	81.54098	1.21'
Interaction	1	221.67066	216.30667	216.30667	3.20'
Within	103	6,954.84220		67.52274	
Total	106	7,360.46729			

* Adjustment term is -5.36399

' Not significant at the 5% level

For the hypothesis H_0', the computed F - ratio was 1.68. Thus, $F_{1,\ 103} = 1.68$ is less than $F_{.95;\ 1,\ 60} = 4.00$ and the hypothesis that the mean of the trigonometry converted gain scores for the control review method was equal to the mean of the trigonometry converted gain scores for the experimental review method was accepted at the five per cent level of significance. Again, it was noted that the value of this F - ratio was much larger than any of the corresponding F - ratios computed for algebra converted gain scores. This result confirmed the previous conclusion that there appeared to be a greater variation between the two review methods for trigonometry achievement than for algebra achievement.

For the hypothesis H_0'', the computed F - ratio was 1.21. Since $F_{1,\ 103} = 1.21$ is less than $F_{.95;\ 1,\ 60} = 4.00$, the hypothesis that the mean of the trigonometry converted gain scores for the above normal-sized class was equal to the mean of the trigonometry converted gain score for the normal-sized class was equal to the mean of the trigonometry converted gain score for the normal-sized classes was accepted at the five per cent level of significance. This confirmed the preceding conclusion based upon the trigonometry posttest scores. Thus, there appeared to be no statistical difference between the type of review method and the sizes of the classes used for this research.

For the hypothesis H_0''', the computed F - ratio was 3.20. Since $F_{1,\ 103} = 3.20$ is less than $F_{.95;\ 1,\ 60} = 4.00$, the hypothesis that there was no interaction between the effect of experimentation and the effect of class size on trigonometry achievement as evidenced by trigonometry converted gain scores was accepted at the five per cent level of significance. This conclusion reinforced the previous conclusion that there was a linear relationship between the variable denoted as experimentation and the variable denoted as class size.

The results of the comparisons reported in this section are summarized in TABLE XLIII and TABLE XLIV. The comparisons revealed no superiority for either review method, either instructor, or either class-size for algebra achievement. Furthermore, these comparisons revealed no superiority for either review method or for class-size for trigonometry achievement. However, these comparisons did reveal a significant statistical interaction between experimentation and instructor. This interaction seemed to favor the writer and the experimental method for achievement in trigonometry. This interaction creates an area into which further investigation could be profitably pursued.

The analyses reported in this chapter reveal no significant statistical difference between review method, class-size, or instructor for either algebra or trigonometry achievement with the exception reported in the preceding paragraph. While some of the comparisons gave F - ratios which appeared to contradict each other, many of these were made from different sub-populations so that the groups were not directly comparable. However, if the main criteria for judging the effectiveness of review method, class-size, or instructor is gain in algebra or trigonometry achievement, the analyses in this chapter revealed no difference in effectiveness except for the combination of instructor and trigonometry.

Chapter V provides an interpretation of these results and analyses through a summary of the findings. Conclusions and implications relative to the findings and recommendations for further research will complete this final chapter.

TABLE XLIII

HYPOTHESES TESTED CONCERNING THE EFFECTS OF EXPERIMENTATION, CLASS-SIZE, AND INSTRUCTOR ON ALGEBRA ACHIEVEMENT

Hypothesis	F - ratio
Equal means of the algebra posttest scores for the control and the experimental review method	0.23'
Equal means of the algebra posttest scores for the pretested and non-pretested groups	0.35'
No interaction between pretesting and experimentation on algebra achievement	1.79'
Equal "corrected" effect on algebra achievement for the control and the experimental method	0.47'
Equal means of the algebra posttest scores for the writer's and Professor X's group	7.72''
No interaction between experimentation and instructor on algebra achievement	9.91''
Equal means of the algebra gain scores for the control and the experimental groups	0.52'
Equal means of the algebra gain scores for the writer's and Professor X's group	2.50'
No interaction between experimentation and instructor on gain in algebra achievement	3.68'
Equal means of the algebra posttest scores for the large-size and the normal-size class	8.17''
No interaction between experimentation and class size on algebra achievement	4.17''
Equal means of the algebra gain scores for large-size and normal-size classes	.54'
No interaction between experimentation and class size on gain in algebra achievement	1.85'

' Not significant at the 5% level

'' Significant at the 5% level

TABLE XLIV

HYPOTHESES TESTED CONCERNING THE EFFECTS OF EXPERIMENTATION, CLASS-SIZE, AND INSTRUCTOR ON TRIGONOMETRY ACHIEVEMENT

Hypothesis	F - ratio
Equal means of the trigonometry posttest scores for the control and the experimental review method	0.43'
Equal means of the trigonometry posttest scores for the pretested and non-pretested groups	0.41'
No interaction between pretesting and experimentation on trigonometry achievement	1.75'
Equal "corrected" effect on trigonometry achievement for the control and the experimental method	0.64'
Equal means of the trigonometry posttest scores for the writer's and Professor X's group	3.08'
No interaction between experimentation and instructor on trigonometry achievement	4.07''
Equal means of the trigonometry gain scores for the control and the experimental group	1.94'
Equal means of the trigonometry gain scores for the writer's and Professor X's group	4.39''
No interaction between experimentation and instructor on gain in trigonometry achievement	8.17''
Equal means of the trigonometry posttest scores for the large-size and the normal-size class	1.11'
No interaction between experimentation and class size on trigonometry achievement	2.24'
Equal means of the trigonometry gain scores for large-size and normal-size classes	1.21'
No interaction between experimentation and class size on gain in trigonometry achievement	3.20'

' Not significant at the 5% level

'' Significant at the 5% level

CHAPTER V

INTERPRETATION OF RESULTS

Review of the Purpose and Design of the Study

This report presents a description of a study that was concerned with an evaluation of a review method for an undergraduate pre-calculus mathematics course featuring a comparatively inexpensive mechanical device. Recognizing that soaring college enrollments and a shortage of qualified professors have created expanding and increasing educational problems, the present study used two different types of review techniques for a pre-calculus course in algebra and trigonometry in an attempt to discover a method for possible alleviation of these problems. The major purpose of the study was to compare the effects of a conventional review method with the effects of an experimental review method, featuring a mechanical device, on students algebraic and trigonometric achievement.

The research design used for this study was a modification of the Solomon Four-Group Design described by Campbell and Stanley[1]. This study was only concerned with the relative merits of the two review methods in producing an increase in algebraic and trigonometric achievement.

---

[1]Donald T. Campbell and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," Handbook of Research on Teaching, ed. N. L. Gage, (Chicago, 1963), p. 175.

Three sections of Mathematics 109, Algebra and Trigonometry, were selected to participate in the present study during the first semester of the 1965-1966 academic year at Wisconsin State University, LaCrosse. Two of the sections were taught by the writer and the other by another member of the mathematics department at Wisconsin State University, LaCrosse. The two sections taught by the writer consisted of fifty-eight students and thirty-six students. The section taught by the other professor consisted of thirty-one students. The student participants were assumed to be a random selection from a normal population of students at Wisconsin State University, LaCrosse. It was established that this was a reasonable assumption.

The instruments used in this study were Cooperative Mathematics Tests of the Educational Testing Service of Princeton, New Jersey. Two tests were given as pretests and two tests were given as posttests. Form A of the Algebra III Test and Form A of the Trigonometry Test were given as pretests and Form B of the Algebra III Test and Form B of the Trigonometry Test were given as posttests. The data thus provided was analysed to determine the relationship between the kind of review method used by the students and their algebraic or trigonometric achievement.

Statistical analyses were made to determine differences between the experimental review method and the control of "conventional" review method. The analysis was primarily concerned with a comparison of the achievement produced by the two different review methods. The design of the study also permitted auxiliary comparisons of the effects of class size and of the class instructor under the two review methods. The statistics employed were Snedecor's F - ratio and chi-square. These statistics were used to compare the sample means and sample variances of the different groups of students involved in the research.

Summary of Findings and Conclusions

Several hypotheses were tested in this research. Since there was a considerable degree of similarity between some of these hypotheses, findings are summarized in terms of the major questions relative to the investigation. The significant findings reported are in terms of the college students included in this study as representative of college students enrolled at Wisconsin State University, LaCrosse.

1. Will the increase in achievement in algebra of the group using the control review method differ from the increase in achievement in algebra of the group using the experimental review method which featured an inexpensive mechanical device consisting of a synchronized slide projector and tape recorder?

There were no significant differences between groups with regard to increase in achievement in algebra. This conclusion was deduced from four separate and distinct analyses of variance and one analysis of covariance. Two of the analyses of variance were computed using means of the algebra posttest scores. The other two analyses of variance were computed using the means of the algebra gain scores. The student's gain score was determined by subtracting the pretest converted score from the posttest converted score. The analysis of covariance was computed by using the algebra posttest converted scores with the algebra pretest converted scores as the covariate. In all of these comparisons, there was no statistical difference between the experimental review method and the control review method for the students involved in this study.

2. Will the increase in achievement in trigonometry of the group using the control review method differ from the increase in achievement in trigonometry of the group using the experimental review method?

There were no significant differences between groups with regard to increase in achievement in trigonometry. Four separate and distinct analyses of variance were used to deduce this conclusion. Two of these were computed using the means of the trigonometry posttest scores. The other two were computed using the means of the trigonometry gain scores. An analysis of covariance was also computed by using the trigonometry posttest converted scores and the trigonometry pretest converted scores. For all of these analyses, there was no statistical difference between the two different review methods for the students involved in this study.

3. Will the increase in achievement in algebra be affected by a combination of class size and review method?

When an analysis of variance was performed on the means of the algebra posttest scores using review method and class size as the sources of variation, there was a significant statistical interaction between the type of review method used and the class size at the five per cent level. The size of the sample used for this analysis was 125. However, since increase in algebra achievement was the desired characteristic, an analysis of variance was performed on the means of the algebra converted gain scores with the same sources of variation. In this analysis, there was no significant statistical interaction between the type of review method used and the size of the class. While these results were not conclusive, the conclusion was made that there was no relationship between the review method used and the size of the class when increase in algebra achievement was the desired characteristic.

This conclusion was limited, of course, to groups of the same size that were used in this research.

4. Will the increase in achievement in trigonometry be affected by a combination of class size and review method?

When an analysis of variance was performed on the means of the trigonometry posttest scores using review method and class size as the sources of variation, there was no significant statistical interaction between the type of review method used and the size of the class. An analysis of variance was also performed on the means of the trigonometry converted gain scores using review method and class size as the sources of variation. There was no significant statistical interaction between the type of review method used and the size of the class. Statistical significance was measured from the five per cent level. The sample size for the first of these analyses was 125 and the sample size of the second analysis was 107, but the second sample was a subset of the first sample. From these results it was concluded that there was no relationship between the review method and the size of the class upon algebraic achievement.

5. Will the increase in achievement in algebra be affected by a combination of different professor and review method?

When an analysis of variance was performed on the means of the algebra posttest scores using review method and different professor as the sources of variation, there was a significant statistical interaction between the type of review method and the particular instructor. This was true at the five per cent level of significance for a sample size of 125. Again, since increase in achievement in algebra was the desired characteristic, an analysis of variance was also performed on the algebra converted gain scores with the same sources of variation.

The sample size for this analysis was also 107. For this analysis, there was no statistical interaction between the type of review method used and the particular instructor. Here again, the results were not completely conclusive, but, since increase in algebra achievement was considered to be the desirable result, the conclusion was made that there was no relationship between the type of review method used and the two instructors who were teaching algebra and who were involved in this research.

6. Will the increase in achievement in trigonometry be affected by a combination of different professor and review method?

There was a significant statistical interaction between the type of review method used and the class instructor, when an analysis of variance was performed on the means of the trigonometry posttest scores using review method and different professor as the sources of variation. When an analysis of variance was performed on the means of the trigonometry converted gain scores using review method and class instructor as the sources of variation, there was also a significant statistical interaction between the review method used and the instructor teaching the class. This was apparent when increase in student achievement in trigonometry was being evaluated. In fact, the F - ratio computed from the converted gain scores was much larger that the F - ratio computed from the posttest scores. This indicated greater interaction between review method and class instructor when measured with converted gain scores than when measured with posttest scores. An examination of the data indicated a much smaller mean for the converted gain scores of the group that was taught by the other professor and that used the control review method than for any of the other groups. There was little difference between the mean of the group that was taught by the writer and that used the experimental review method and the mean of the group that was taught by the writer and that used the control method. Both of these groups has means which were nearly twice the value of the smallest mean mentioned above. The conclusion was then made that the increase in trigonometry achievement was affected by a combination of review method and professor. The data seemed to suggest that the combination of control review method and Professor X was least effective in producing an increase in student achievement in trigonometry. This certainly suggests that this is an area where further research would be desirable and profitable.

Implications

The findings suggest the following implications for the further study of this type of mechanized review device:

1. Information obtained in the present study suggests the utility of the research design in further study of the use of this mechanized review method by undergraduate students of pre-calculus mathematics. Since differences in achievement between groups were primarily concerned with an increase or gain in achievement, factors in the students' background that were not included in this investigation may have influenced his increase in achievement. More specific differences would be noted when considering this increase in achievement as an individual rather than a group phenomenon.

2. Individuals investigating the use of the review method discussed in this study should construct review materials which will reflect the experience and attitudes of all persons involved in the study. This should reduce the interaction between type of review method and class instructor. Perhaps the only way to minimize this interaction will be to immerse each individual involved as completely as possible in the study itself.

3. The findings reveal that the differences in achievement in algebra were not affected as much by combinations of review method, class size, and instructor as was achievement in trigonometry. This suggests that there may be something inherent in the course content in trigonometry that is influenced by different combinations of these factors. The results of the study seem to imply that the greatest increase in achievement is brought about by a combination of trigonometry content, mechanized review method, and a large class.

Suggestions for Further Study

The conclusions and implications of the present study suggest more refined and intensive investigations which will consider the following recommendations:

1. An intensive analysis should be employed to consider the type of student that would have the greatest increase in achievement using this review method for algebraic content and for trigonometric content. This would permit a better

evaluation of the optimal use for this mechanized review method.

2. It appears to be advisable to investigate in much more depth the relationship between the use of mechanized review method and the class instructor. That is, what kind of professor is most effective in producing increase in achievement when using this review method?

3. It also appears to be advisable to investigate much more deeply the relationship between the use of this review method and the size of the class involved.

4. The multiple relationship between this review method, the particular instructor, and the size of the class should be investigated to determine the combination that will deliver the optimum results in terms of increase in achievement.

Concluding Remarks

In the past decade, much has been written about mathematics education in this country. Initially, most of this discussion was focused upon the elementary and secondary mathematics curriculum. In the last few years, however, the collegiate mathematics program has also been subjected to a great amount of penetrating and enlightening review. Some of the impetus for this review has been given by the soaring enrollments in higher education, the shortage of qualified mathematics professors, and the tremendous increase in technology and knowledge. These three events have been well documented, and, if our system of higher education in mathematics is to remain dynamic and continue maximally to contribute to our society and culture, then methods must be found that will enable higher education effectively and economically to solve these problems.

This study has been an attempt to secure some evidence that may enable higher education, and particularly higher education in mathematics, to solve the problem of soaring enrollments and a growing shortage of qualified mathematics professors. This research has indicated a method that might be used to relieve some of the pressure on understaffed collegiate mathematics departments. The findings suggest the possibility of handling large sections of pre-calculus algebra and trigonometry with the aid of the mechanized review method described in this study. While this research was not adequate for a final and definitive statement concerning this possibility, the findings were adequate enough to suggest that further study in this area would be desirable and that such study should be heartily encouraged.

SELECTED BIBLIOGRAPHY

Baskin, Samuel. "Independent Study: Methods, Programs, and for Whom?," Higher Education in an Age of Revolutions, Ed. G. Kerry Smith. Washington, D. C., Association for Higher Education, 1962, 65-68.

Belcastro, Frank P. "Programmed Learning and Intelligence," School Science and Mathematics, LXVI (January, 1966), 29-36.

Benner, Charles P. and Curtis A. Rogers. "A New Plan for Instructing Large Classes in Mathematics by Television and Films," The Mathematics Teacher, LIII (May, 1960), 371-375.

Brearley, H. C. "College Classroom Teaching: Problems and Procedures," Peabody Journal of Education, XXXVII (Summer, 1959), 66-76.

Brown, James W. and James W. Thornton, Jr. College Teaching: Perspectives Guidelines, New York: McGraw-Hill Book Company, IUc., 1963.

Brown, Robert S. "Survey of Ohio College Opinions with Reference to High School Mathematics Program," The Mathematics Teacher, LIV (April, 1963), 245-247.

Bruner, Jerome S. The Process of Education, Cambridge: Harvard University Press, 1961.

Burkhard, Sarah. "A Study of Concept Learning in Differential Calculus," Dissertation Abstracts, XVI, No. 2 (Columbia University, 1956).

Butler, Charles H. and F. Lynwood Wren. Teaching of Secondary Mathematics, New York: McGraw-Hill Book Company, Inc., 1941.

Campbell, Donald T. and Julian C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching" Handbook of Research on Teaching, Ed. N. L. Gage. Chicago: Rand McNally and Company, 1963, 171-246

Carpenter, C. R. "Teaching and Learning by Television," The Two Ends of the Log; Teaching and Learning in Today's Colleges. Ed. Russell M. Cooper. Minneapolis: University of Minnesota Press, 1958, 209-221.

Churchill, R. and John P. Churchill. "Conservation of Teaching Time Through the Use of Lecture Classes and Teaching Assistants," Journal of Educational Psychology, IL (December, 1958), 324-327.

Dale, Edgar. "New Techniques of Teaching," The Two Ends of the Log; Learning and Teaching in Today's Colleges, Ed. Russell M. Cooper. Minneapolis: University of Minnesota Press, 1958, 191-208.

__________, Wisconsin State Universities Report, XV (January, 1966), 3.

Dubisch, Roy. The Teaching of Mathematics. New York and London: John Wiley and Sons, Inc., 1963.

__________. "Individual Differences in College Mathematics," School and Society, LXXXIV (December 8, 1956), 204.

Fehr, Howard F., ed., The Learning of Mathematics; Its Theory and Practice. Washington, D. C.: The National Council of Teachers of Mathematics, 1953.

Gage, N. L. "Paradigms for Research on Teaching," Handbook of Research on Teaching. Ed. N. L. Gage. Chicago: Rand McNally and Company, 1963, 94-141.

Guenther, William C. Analysis of Variance, Englewood Cliffs, N. J.: Prentice - Hall, Inc., 1964.

Guenther, William C. Concepts of Statistical Inference. New York: McGraw - Hill Book Company, Inc., 1965.

Hawkins, Thomas E. "A Volunteer Tutorial System," Phi Delta Kappan, XL (January, 1959), 168-169.

Highet, Gilbert. The Art of Teaching. New York: Vintage Books, 1950.

Hildebrandt, E. H. C. "For a Better Mathematics Program in the College," The Mathematics Teacher, IL (February, 1956), 89-99.

Hyman, Lawrence W. "Advancing Education by Eliminating Classes," Journal of Higher Education, XXXII (April, 1961), 213-215.

Johnson, Palmer O. and Robert W. B. Jackson. Modern Statistical Methods: Descriptive and Inductive. Chicago: Rand McNally and Company, 1959.

Jones, P. S. "Content in the Teaching of Mathematics," Virginia Journal of Education, LIII (May, 1960), 15-21.

Kemeny, John G. "Report to the International Congress of Mathematicians," The Mathematics Teacher, LVI (February, 1963), 66-78.

Kemeny, J. G. "Rigor Vs Intuition in Mathematics," The Mathematics Teacher, LIV (February, 1961), 66-74.

Kinsella, J. J. "Is Research in Mathematics Education Really Necessary?," The Mathematics Teacher, L (April, 1957), 300-301.

Kuuisto, Allan A. "What are the Most Effective Methods of Dealing with Larger Numbers of Students?," Higher Education in an Age of Revolutions. Ed. G. Kerry Smith. Washington, D. C.: Association for Higher Education, 1962, 173-176.

Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Mifflin Company, 1953, 40.

Lumsdaine, A. A. "Instruments and Media of Instruction," Handbook of Research on Teaching. Ed. N. L. Gage. Chicago: Rand McNally and Company, 1963, 583-682.

Mallinson, G. G. "The Five Most Needed Research Investigations in the Teaching of Science and Mathematics," School Science and Mathematics, LIV (June, 1954), 428-430.

Marr, J. M. and Others. "Contributions of the Lecture to College Teaching," Journal of Educational Psychology, LI (October, 1960), 277-284.

Mayer, Martin. The Schools. New York: Harper and Brother, 1961.

McKeachie, W. J. "Procedures and Techniques of Teaching: A Survey of Experimental Studies," The American College: A Psychological and Social Interpretation of the Higher Learning. Ed. Nevitt Sanford. New York: John Wiley and Sons, Inc., 1962, 312-356.

McKeachie, W. J. "Research on Teaching at the College and University Level," Handbook of Research on Teaching. Ed. N. L. Gage. Chicago: Rand McNally and Company, 1963, 1118-1172.

Mood, Alexander M. and Granklin A. Graybill. Introduction to The Theory of Statistics. New York: Mc Graw - Hill Book Company, Inc., 1963.

Pressey, Sidney L. "Educational Acceleration: Occasional Procedure or Major Issue," The Personnel and Guidance Journal, (September, 1962), 12-16.

Price, G. Baley. "New Perspectives on Teaching Mathematics," Higher Education in an Age of Revolutions. Ed. G. Kerry Smith. Washington, D. C.: Association for Higher Education, 1962, 100-102.

Price, H. V. "Needed Research in the Teaching of Science and Mathematics," School Science and Mathematics, LVII (May, 1957), 389-390.

Roberson, J. A. "Teaching Effectiveness and Class Size," Journal of Engineering Education, XXX (1959), 401-420.

Rosenbloom, Paul C. "The Role of Mathematics and Science in a General Education," mimeographed article, (Minneapolis, 1959), 1.

Sanford, Nevitt, Ed., The American College: A Psychological and Social Interpretation of the Higher Learning. New York: John Wiley and Sons, Inc., 1962.

Simpson, T. M. "Mathematics in the College General Education Program," The Mathematics Teacher, L (February, 1957), 155-159.

Smith, Roland Frederick. "An Experimental Comparison of Two Liberal Arts Courses in General Mathematics at Syracuse University," Dissertation Abstracts, XV, No. 6 (Syracuse University, 1955).

Sueltz, Ben A. "These Things We Believe," The Mathematics Teacher, IL (January, 1956), 19-21.

Trippet, Byron K. "Are Fundamental Changes Required in Higher Education?," Goals for Higher Education in a Decade of Decision. Ed. G. Kerry Smith. Washington, D. C.: Association for Higher Education, 1961, 24-29.

Van Dalen, Debold B. and William J. Meyer. Understanding Educational Research. New York: McGraw - Hill Book Company, Inc., 1962.

Wert, James E., Charles O. Neidt, and J. Stanley Ahmann. Statistical Methods in Educational and Psychological Research. New York: Appleton - Century - Crofts, Inc., 1954.

Williams, Horace E. "A Study of the Effectiveness of Classroom Teaching Techniques Following a Closed - Circuit Television Presentation in Mathematics," The Mathematics Teacher, LIV (February, 1963), 94-97.

Williamson, Robert Gordon. "A Theory of Learning and Its Application to a Class in College Mathematics," Dissertation Abstracts, XVI, No. 4 (University of Maryland, 1956).

APPENDIX

C_1 TEST RESULTS

Student Number	Trigonometry Scores				Algebra Scores			
	Pretest		Posttest		Pretest		Posttest	
	Converted	Raw	Converted	Raw	Converted	Raw	Converted	Raw
1	149	14	151	16	138	9	144	15
2	147	13	173	31	167	28	170	31
3	150	15	163	24	144	13	152	20
4	140	9	154	18	144	13	142	14
5	144	11	154	18	157	21	157	23
6	137	7	136	6	136	8	123	2
7	149	14	145	12	146	14	144	15
8	144	11	164	25	140	10	162	26
9	132	4	145	12	138	9	149	18
10	139	8	137	7	141	11	139	12
11	137	7	134	5	136	8	147	17
12	155	18	155	19	143	12	156	22
13	145	12	152	17	147	15	149	18
14	145	12	149	15	141	11	165	28
15	142	10	155	19	140	10	159	24
16	154	17	165	26	149	16	167	29
17	144	11	146	13	136	8	146	16

APPENDIX (Continued)

E_1 TEST RESULTS

Student Number	Trigonometry Scores				Algebra Scores			
	Pretest		Posttest		Pretest		Posttest	
	Converted	Raw	Converted	Raw	Converted	Raw	Converted	Raw
1	140	9	160	22	141	11	157	23
2	137	7	168	28	147	15	156	22
3	134	5	157	20	147	15	164	27
4	134	5	143	11	135	7	154	21
5	155	18	164	25	155	20	161	25
6	142	10	160	22	144	13	162	26
7	145	12	148	14	150	17	156	22
8	140	9	152	17	143	12	154	21
9	135	6	154	18	136	8	154	21
10	129	2	136	6	136	8	144	15
11	144	11	157	20	136	8	146	16
12	132	4	146	13	146	14	142	14
13	145	12	157	20	158	22	164	27
14	142	10	139	8	138	9	144	15

APPENDIX (Continued)

C_2 TEST RESULTS

Student Number	Trigonometry Scores				Algebra Scores			
	Pretest		Posttest		Pretest		Posttest	
	Converted	Raw	Converted	Raw	Converted	Raw	Converted	Raw
1	147	13	164	25	158	22	161	25
2	154	17	170	29	153	19	156	22
3	142	10	143	11	138	9	156	22
4	131	3	149	15	133	6	146	16
5	137	7	154	18	138	9	162	26
6	149	14	161	23	155	20	162	26
7	134	5	157	20	153	19	157	23
8	145	12	152	17	149	16	170	31
9	126	0	149	15	141	11	147	17
10	147	13	154	18	140	10	156	22
11	147	13	149	15	147	15	156	22
12	152	16	152	17	147	15	166	22
13	140	9	163	24	143	12	161	25
14	129	2	149	15	143	12	151	19
15	154	17	149	15	135	7	154	21
16	149	14	167	27	164	26	182	38
17	139	8	154	18	152	18	156	22
18	132	4	151	16	138	9	154	21
19	150	15	167	27	158	22	164	27
20	129	2	145	12	138	9	154	21
21	142	10	161	23	155	20	169	30

APPENDIX (Continued)

C_2 TEST RESULTS

Student Number	Trigonometry Scores				Algebra Scores			
	Pretest		Posttest		Pretest		Posttest	
	Converted	Raw	Converted	Raw	Converted	Raw	Converted	Raw
22	152	16	173	31	155	20	164	27
23	135	6	148	14	136	8	146	16
24	150	15	165	26	160	23	165	28
25	150	15	152	17	136	8	147	17
26	142	10	145	12	143	12	154	21
27	149	14	154	18	147	15	157	23
28	135	6	143	11	135	7	146	16
29	144	11	164	25	161	24	164	27

$C_{4,P}$ TEST RESULTS

Student Number	Trigonometry Scores				Algebra Scores			
	Pretest		Posttest		Pretest		Posttest	
	Converted	Raw	Converted	Raw	Converted	Raw	Converted	Raw
1	149	14	151	16	141	11	151	19
2	149	14	154	18	150	17	157	23
3	139	8	165	26	146	14	167	29
4	149	14	171	30	146	14	161	25
5	145	12	170	29	153	19	154	21
6	144	11	161	23	144	13	159	24
7	135	6	142	10	141	11	147	17
8	162	22	173	31	164	26	172	32
9	144	11	163	24	140	10	159	24

APPENDIX (Continued)

E_2 TEST RESULTS

Student Number	Trigonometry Scores				Algebra Scores			
	Pretest		Posttest		Pretest		Posttest	
	Converted	Raw	Converted	Raw	Converted	Raw	Converted	Raw
1	135	6	142	10	138	9	142	14
2	152	16	165	26	150	17	162	26
3	142	10	167	27	150	17	164	27
4	140	9	145	12	140	10	159	24
5	150	15	164	25	158	22	169	30
6	132	4	148	14	150	17	149	18
7	152	16	154	18	149	16	154	21
8	132	4	155	19	133	6	147	17
9	137	7	160	22	140	10	157	23
10	145	12	163	24	147	15	162	26
11	147	13	151	16	150	17	157	23
12	149	14	155	19	147	15	156	22
13	150	15	155	19	141	11	159	24
14	142	10	143	11	146	14	161	25
15	149	14	165	26	164	26	159	24
16	147	13	152	17	150	17	156	22
17	145	12	155	19	146	14	152	20
18	139	8	155	19	143	12	159	24
19	131	3	145	12	146	14	161	25
20	131	3	157	20	140	10	161	25
21	127	1	151	16	135	7	156	22

APPENDIX (Continued)

E_2 TEST RESULTS

Student Number	Trigonometry Scores				Algebra Scores			
	Pretest		Posttest		Pretest		Posttest	
	Converted	Raw	Converted	Raw	Converted	Raw	Converted	Raw
22	144	11	167	27	153	19	159	24
23	139	8	154	18	147	15	157	23
24	165	24	176	33	164	26	162	26
25	139	8	154	18	141	11	151	19
26	142	10	173	31	155	20	167	29
27	132	4	145	12	147	15	154	21
28	155	18	157	20	146	14	165	28
29	137	7	149	15	141	11	151	19

$E_{4,P}$ TEST RESULTS

Student Number	Trigonometry Scores				Algebra Scores			
	Pretest		Posttest		Pretest		Posttest	
	Converted	Raw	Converted	Raw	Converted	Raw	Converted	Raw
1	149	14	164	25	141	11	167	29
2	144	11	154	18	143	12	152	20
3	145	12	155	19	141	11	152	20
4	139	8	154	18	146	14	157	23
5	145	12	157	20	136	8	147	17
6	137	7	157	20	140	10	154	21
7	137	7	139	8	136	8	152	20
8	127	1	145	12	136	8	139	12
9	142	10	161	23	147	15	157	23

APPENDIX (Continued)

$C_{4,N}$ TEST RESULTS

Student Number	Trigonometry Scores Posttest Converted	Raw	Algebra Scores Posttest Converted	Raw
1	160	22	154	21
2	160	22	164	27
3	160	22	152	20
4	160	27	159	24
5	148	14	149	18
6	149	15	164	27
7	167	27	167	29
8	149	15	149	18
9	146	13	169	30
10	152	17	165	28

$E_{4,N}$ TEST RESULTS

Student Number	Trigonometry Scores Posttest Converted	Raw	Algebra Scores Posttest Converted	Raw
1	165	26	162	26
2	152	17	165	28
3	145	12	149	18
4	170	29	162	26
5	143	11	144	15
6	160	22	164	27
7	133	4	147	17
8	136	6	141	13

VITA

Roland Walter Christensen

Candidate for the Degree of

Doctor of Education

Thesis: A COMPARISON OF TWO REVIEW METHODS FOR ALGEBRA AND TRIGONOMETRY AT WISCONSIN STATE UNIVERSITY, LACROSSE

Major Field: Higher Education

Biographical:

Personal Data: Born near Soldier, Iowa, July 13, 1927, the son of Harry N. and Elvira C. Christensen.

Education: Attended elementary school near Ute, Iowa, and near Burt, Iowa; graduated from Burt High School, Burt, Iowa, in 1945; attended Southwestern University, Georgetown, Texas, in 1945-1946; received the Bachelor of Arts degree from Iowa State Teachers College, Cedar Falls, Iowa, in August, 1950; attended Brigham Young University in the summer of 1954; attended the University of Illinois in the summers of 1957 and 1958; attended the University of Minnesota as a member of a National Science Foundation Academic Year Institute from June, 1959 to August, 1960; attended the University of Minnesota during the summer of 1961; attended Oklahoma State University during the summer of 1963 and from June, 1964 to August, 1965; completed requirements for the Doctor of Education degree in May, 1967.

Professional experience: High school teacher of Mathematics and General Science in Franklin Consolidated School, Latimer, Iowa, 1950-1951; high school teacher of Mathematics and English in Crystal Lake Consolidated School, Crystal Lake, Iowa, 1951-1952; high school teacher of General Science in Estherville Consolidated School, Estherville, Iowa, 1952-1953; high school teacher of Mathematics in Vandalia Community Unit School, Vandalia, Illinois, 1953-1957; high school teacher of Mathematics in Streator Independent School, Streator, Illinois, 1957-1959; college teacher of Mathematics in Wisconsin State College, now Wisconsin State University, LaCrosse, Wisconsin, 1960-1967.